

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 08:18:22 ; Search time 10398.6 Seconds

(without alignments)  
10745.552 Million cell updates/sec

Title: US-10-603-260-1

Perfect score: 2578  
Sequence: 1 agcttgacattacagcca.....tgataataatggtttcttag 2578

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_da:\*
- 2: gb\_hg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vt:\*
- 15: em\_da:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pac:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vt:\*
- 30: em\_hg\_hum:\*
- 31: em\_hg\_in:\*
- 32: em\_hg\_other:\*
- 33: em\_hg\_mus:\*
- 34: em\_hg\_pln:\*
- 35: em\_hg\_rod:\*
- 36: em\_hg\_mam:\*
- 37: em\_hg\_vit:\*
- 38: em\_sy:\*
- 39: em\_hgo\_hum:\*
- 40: em\_hgo\_mus:\*
- 41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	574	22.3	10977	1	AE004352	AE004352 Vibrio ch
C 2	467	18.1	306147	1	AE016805	AE016805 Vibrio vu
C 3	462.2	17.9	241900	1	AP005337	AP005337 Vibrio vu
C 4	403	15.6	303450	1	AP005085	AP005085 Vibrio pa
C 5	230.4	8.9	247450	1	AP005349	AP005349 Vibrio vu
C 6	230.4	8.9	302331	1	AE016810	AE016810 Vibrio vu
C 7	226.4	8.8	4228	6	AX023555	AX023555 Sequence
C 8	226.4	8.8	5391	6	AX023554	AX023554 Sequence
C 9	226.4	8.8	7057	6	AX023552	AX023552 Sequence
C 10	226.4	8.8	9402	6	AX023551	AX023551 Sequence
C 11	226.4	8.8	10890	6	AX023553	AX023553 Sequence
C 12	224.8	8.7	1073	6	AR371489	AR371489 Sequence
C 13	224.8	8.7	1073	6	AX195443	AX195443 Sequence
C 14	224.8	8.7	2297	6	BD261822	BD261822 Recombina
C 15	224.8	8.7	2397	6	AX027819	AX027819 Sequence
C 16	224.8	8.7	2368	12	AF264618	AF264618 Cloning v
C 17	224.8	8.7	2417	12	CVSPSP70	CVSPSP70 Cloning v
C 18	224.8	8.7	2419	12	CVSPSP71	CVSPSP71 Cloning vec
C 19	224.8	8.7	2450	12	CYT74374	CYT74374 Cloning vec
C 20	224.8	8.7	2462	6	BD002331	BD002331 Transgeni
C 21	224.8	8.7	2462	12	CVSPSP72	CVSPSP72 Cloning vec
C 22	224.8	8.7	2464	12	CVSPSP73	CVSPSP73 Cloning vec
C 23	224.8	8.7	2512	12	ASA2679	ASA2679 Synthetic
C 24	224.8	8.7	2580	6	AR199681	AR199681 Sequence
C 25	224.8	8.7	2655	12	SYNPU88V	SYNPU88V Cloning
C 26	224.8	8.7	2655	12	SYNPU89V	SYNPU89V Cloning
C 27	224.8	8.7	2668	12	SYNPU90V	SYNPU90V Cloning
C 28	224.8	8.7	2674	12	SYNPU91V	SYNPU91V Cloning
C 29	224.8	8.7	2680	12	SYNPU92V	SYNPU92V Cloning
C 30	224.8	8.7	2680	12	SYNPU93V	SYNPU93V Cloning
C 31	224.8	8.7	2680	12	SYNPU94V	SYNPU94V Cloning
C 32	224.8	8.7	2686	6	AO27710	AO27710 Cloning vec
C 33	224.8	8.7	2686	6	AX743996	AX743996 Sequence
C 34	224.8	8.7	2686	6	BD000457	BD000457 Sequence
C 35	224.8	8.7	2686	12	SYNPU918V	SYNPU918V Cloning
C 36	224.8	8.7	2686	12	SYNPU919V	SYNPU919V Cloning
C 37	224.8	8.7	2695	12	SYNPU920V	SYNPU920V Cloning
C 38	224.8	8.7	2695	12	SYNPU921V	SYNPU921V Cloning
C 39	224.8	8.7	2695	12	SYNPU922V	SYNPU922V Cloning
C 40	224.8	8.7	2695	12	SYNPU923V	SYNPU923V Cloning
C 41	224.8	8.7	2695	12	SYNPU924V	SYNPU924V Cloning
C 42	224.8	8.7	2701	12	SYNPU925V	SYNPU925V Cloning
C 43	224.8	8.7	2704	12	SYNPU926V	SYNPU926V Cloning
C 44	224.8	8.7	2704	12	SYNPU927V	SYNPU927V Cloning
C 45	224.8	8.7	2710	12	SYNPU928V	SYNPU928V Cloning

# ALIGNMENTS

RESULT 1  
AE004352/c 10977 bp DNA linear BCT 10-OCT-2003  
LOCUS Vibrio cholerae O1 biovar eltor str. N16961 chromosome II, section 9 of 9 of the complete chromosome.  
ACCESSION AE004352 AE003853  
VERSION AE004352.1 GI:9657475  
KEYWORDS  
SOURCE  
ORGANISM  
Vibrio cholerae O1 biovar eltor str. N16961  
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
Vibrionaceae; Vibrio.  
1 (bases 1 to 10977)  
Heidelberg, U.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,



CDS 6518..7825  
 /gene="UCA0098"  
 /note="similar to GB:J05568 SP:P18133 PID:147307 GB:U00096  
 PID:1651456; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="nicotinate phosphoribosyltransferase"  
 /protein\_id="AA096012.1"  
 /db\_xref="GI:9657484"  
 /translation="MNPRLSPHIIIRSLDDAYKIMQAIIHFYPDVSYELIYR  
 SEDASGLDAIRQELIAGLTFESDADIIHLTQAPLKLATPQSLAFYFVQEOY  
 KMDIYQGGKQQLIRISRWMDIILYETLVAIVSEYRSQRAEAPADIPLYLTV  
 EKVQKAEIERGINNPSLTEMGTRRPSQVORVLAELKQEIIPQWLTGTSNHPAR  
 BBDKLEIGTIAHEMFHGOALVNERDSQVORVLAELKQEIIPQWLTGTSNHPAR  
 DNPRLIANYGVNRDSCSPKRMKNAIHIQQLIDPTTCLTFPSGLDLPDQLELC  
 EYFAGRKISFQIGFTINDLANNRANRAGVIRPUSIVTKIAECQGRVAKISQPEK  
 AMCEDPILANIKRRFNIELVDALIQELRHQKSPHYISAA"  
 7856..8896  
 /gene="VCA0099"  
 7856..8896  
 /gene="VCA0099"  
 /note="similar to GB:U00096 PID:1742678 PID:1742686  
 PID:1742694 PID:1742718; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="oxidoreductase, Gfo/Idh/Moa family"  
 /protein\_id="AA096013.1"  
 /db\_xref="GI:9657485"  
 /translation="MNPRLKVGIVIGYSYAKTFLPDIPLTSLSSSQQVY  
 EDWQPIAFDSQRLITSDVDLITAPNNVHFPPLAKLEHGVKIVYKRPVTOI  
 EGRTEILARQGLILSVFNHRWDGFLTVKLIIEGQGLDVGKIVESHEDRVPVY  
 RQRMEOAEGGGILFPLDAPHLIDQALFGLPQSLSESDGKMDPATITDYLQY  
 YPOHVLRIANLISPEPNRYOVLGSLKYKGLDPEDEMLKXGEPTHQWQENP  
 ROYGLTXAEGENVITELGGYQHTFTQVVAIRNGASNPASALQSLITALE  
 SSAQQLAVTL"  
 complement (8933..9304)  
 /gene="VCA0100"  
 complement (8933..9304)  
 /gene="VCA0100"  
 /note="similar to GP:7226016; identified by sequence  
 22.3%; Score 574; DB 1; Length 10977;  
 Query Match  
 Best Local Similarity 64.3%; Pred. No. 5,4e-157;  
 Matches 1028; Conservative 0; Mismatches 440; Indels 130; Gaps 6;  
 gene  
 CDS  
 23 ACTGCATCGTACTCGCGCGGCACTTGTGCCAGTGGCGCGCTACTGATTCAAG 82  
 6987 ACCTTCAGGTAAATCAAGCAGGCACTCAAGCCCAAGCTGACGGCTCGTATTCAGAT 6928  
 83 ATTAAAGCCATGACTAGCGGTTTCATATAAATGATGTCGTGCCACGTAACCTTGAATGCG 142  
 6927 ACGATCGCATGACTAAGGTTTCATACAAATCGGTCAAGCACTCCAGAAATCTG 6868  
 143 ATAGCAGGTGGCGTTTCCCTTGTCTGAGATCCCGATTCAATTTGGCGAATCGGGT 202  
 6867 ATCCGCAAGTGTCTTTCCTCGCTGTGTGCAATCCCATTTCTACTGTCTTGAAGC 6808  
 203 TGAATATGAAATACGTAATGACTGTAAATAAAGTACGATTCMAATGAGTGCATGCTGC 262  
 6807 ACAAAATGGAATTAAGCAAGATTGCAAGAAAGTCCGCTTGAGATGAGGCAATGCTGA 6748  
 263 TCTAATAAACAATGTCCGATCCGAAAAGGCAATGAAGCAACTGATTTATTTCTTGG 322  
 6747 GTCAATATATGAAATGCGCAATCCGAAAAGGCAATGCGCAAGATGACTATCTTTGA 6688  
 323 CGTACTTCCTCTAATAATCGCTATGCTTCATCACTGCGCAATCAATCAATTCATAGGC 382  
 6687 CGGATGCGATGAAGAAGTCCGCTGCGCATCTCTTCACATGCGTAACTAATTCATAGGC 6628  
 383 ACCTCAACATCCGATTAACAAGATGAACGGCTGCGATCATATTGATTTTATAGCATCA 442  
 6627 ACCGACATCCGATTAATAAATGATGATCGCTTGCAATCATATTGATTTTATAGCATCA 6568

QY 443 AGATCCAAATAAATCGCGGATTAAGAGAGAAATAAGCGATCGCTCATGATGATGCCA 502  
 DB 6567 AGATCGAGCAGGCTACCAATGATGCGGAGAGAAACGCGGTGATTCATGAT--TGCCG 6510  
 QY 503 TCCCTTCGTCGCGTTTCATCAATCAATTAAGTAAAGCAAGTGTGCTAAC--TTGGG 561  
 DB 6509 TCCCTTCGTCACATTCGCGAT-----GTTAGAACATTTTGTTCATCTTTTGA 6460  
 QY 562 CGAACATTAAGTACCTTGTGAAGTTTGTCACTTTTGTGCAACCTA----- 610  
 DB 6459 GGCACGATATGAGCCCTCTATGTTGTCACTTTTGTGCAACCAACATTAAGAGA 6400  
 QY 611 ----- 610  
 DB 6399 AGGAATCTCGTTAAGACAGCAGCAGCCTTCTAATCCAAACCTGCTGAGTTGC 6340  
 QY 611 -----GTCAGTCGTTATTTGGCCTTATTAATATATGATAT 647  
 DB 6339 AGGTAAGCGGCAATAGTGAATCCCATGAGCATAGTACAGATGTATGCGGTGAG 6280  
 QY 648 TGAAGGGTAAAGACGATGATTAACAATTA-----CAGTA 685  
 DB 6279 TGAATGCTGCACACACCGCTGACGCTTCAGTAGAGAGGATATGATTCAGTGTCT 6220  
 QY 686 CTCTGTATCTGATGATTTGTTGTCACAAAGCTTATTTACATTTGACCATCATGTC 745  
 DB 6219 ATAACTGCCCAATTCAGTAAATCAATTTGTTATTTACATTTGACCTTACATTC 6160  
 QY 746 ACTTACTTAATAATTAAGCCGCTGTTATTAAGGAGCACTTATGATGTCATATGATA 805  
 DB 6159 AGTTACTTAATAA--ACCGCACCTTATATGAGAACCTTGTATGATGTCACCATTTGATA 6101  
 QY 806 TGAATTTGTCGCGCTTGGCGCGAATCTATCCAGGTTTACGTGGAACGCTTAATC 865  
 DB 6100 TGAATCTTTGCGCTTGGCGCGAATCTATCCAGGTTTACGTGGAACGCTTAATC 6041  
 QY 866 CAATCGGCGCAATGTTGTTAAATGCGCATTTGCTGGCGGATATGTTAGCAAGATA 925  
 DB 6040 CAATCGGCGCTTAATGCGGCTTTGCGGCAATTCGCGTGGTGGGTGTTGATGAGGATC 5981  
 QY 926 TGAACGCTCATGTTGAGAGAACTGTTGATGAGATTTTATGAGAGAGACGATTT 985  
 DB 5980 TCAAGCGCAAGGTGCGAACCCTGTCACAAATTTATTTTCGCGCGTGGCGCATCT 5921  
 QY 986 GTGCGAAAAGTCCATTAATCTTAATTTATCAGCATCCGCTGTGATGAGCAACC 1045  
 DB 5920 GCGGCAAGAAATCCATTAATTTATCCCACTTATCATGATATCCTTTGGAGAGGATATC 5861  
 QY 1046 CCAAGCGCATCCGAATGTTGAGATGATATTTTCCGATTAACGCTTTATTAACCCGT 1105  
 DB 5860 CTAAAGCGCATCCGAAGGTTGAGATGATATTTCTCATATGCTGCTCAATCAAT 5801  
 QY 1106 GGAATGCAAAACAATAGAAATTTGATGACACCCGAGCGCGCTAATGTTGATGATC 1165  
 DB 5800 CCAAGCTCAAGCATTTGAAGAGGCTGAATGATTAAGCCGTTAATGTTGATGATC 5741  
 QY 1166 TTCTATCTTAATCAAGAAAGAAATGCGCTGCTTTGATCATGTCGCGCAATTCAGC 1225  
 DB 5740 TGAAGAGATTTTGAAGGTAACAATGTTCTGCGTTGATCATGTTGCGCAATTCAGC 5681  
 QY 1226 ATGCGTGGCAAAATTAAGCGCTGCGTGAATATCAATCCGTTGATCAATTTTCACTAG 1285  
 DB 5680 ATGCAATGAAATAAATCGCGCGAGCGGTGAGTACCTTGTAGTGTCTGCTTCACTAG 5621  
 QY 1286 AAAAAGTTTTTATGAGCGGATATTAATGATGCTTACGCAATTTGCGCTGAAAGTTA 1345  
 DB 5620 AGCTGAGTTTTTGTGGCGGACATTAATGAGCTTACCCAAATTTGAGATGAGAGTGA 5561  
 QY 1346 ATGCGATGACATTAAGCGCGCTTGAATCAATTAACGGGTTGATGCTGATCAATTA 1405  
 DB 5560 ATGCGATGAGCGGTAATAAGCTGCTGATGATACCGGGGTAATTTGTCAGTACCAATTA 5501

QY 1406 TGCGCGCATCTTGTAAAGGCAAGGACCAACGACCGTATGCTTATGCGCAGTC 1465  
DB 5500 TGGCCCTCTAGCAAGGTAAAGGCGCAACCTGCGAGCGTTATGCTTAGCAGTA 5441  
QY 1466 ATGAAGTCACTATTTTCAACCTGTTTACGAGTTAA 1503  
DB 5440 ACGAAGTCATCTATTTTACAGACTTGCTTGCAGGCTAA 5403

RESULT 2  
AE016805 306147 bp DNA linear BCT 29-SEP-2003  
LOCUS Vibrio vulnificus CMCP6 chromosome I section 9 of 11 of the  
DEFINITION complete sequence.  
ACCESSION AE016805 AE016795  
VERSION AE016805.1 GI:27361837  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS Kim,Y.R., Lee,S.E., Kim,C.M., Kim,S.Y., Shin,E.K., Shin,D.H.,  
Chung,S.S., Choy,H.E., Progulskoe-Fox,A., Hillman,D., Handfield,M.  
and Rhee,J.H.  
TITLE Characterization and Pathogenic Significance of Vibrio vulnificus  
Antigens Preferentially Expressed in Septicemic Patients  
JOURNAL Infect. Immun. 71 (10), 5461-5471 (2003)  
PUBMED 14500463  
REFERENCE 2 (bases 1 to 306147)  
AUTHORS Jeong,H., Moon,Y.H. and Kim,J.J.  
TITLE Direct Submision  
JOURNAL Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong,  
Yuseong-gu, Daejeon 305-811, South Korea  
REFERENCE 3 (bases 1 to 306147)  
AUTHORS Rhee,J.H., Kim,S.Y., Chung,S.S., Lee,S.E. and Choy,H.E.  
TITLE Direct Submision  
JOURNAL Submitted (13-DEC-2002) Department of Microbiology, Genome Research  
Center for Enteropathogenic Bacteria, Chonnam National University  
Medical School, Hak-1-Dong, Dong-gu, Kwang-Ju 501-746, South Korea  
FEATURES  
source  
1. 306147  
/organism="Vibrio vulnificus CMCP6"  
/mol\_type="genomic DNA"  
/strain="CMCP6"  
/db\_xref="taxon:216895"  
/chromosome="I"  
92..1075  
/locus\_tag="VV12370"  
92..1075  
/locus\_tag="VV12370"  
/note="COG0016"  
/codon\_start=1  
/transl\_table=11  
/product="Phenylalanyl-tRNA synthetase, alpha chain"  
/protein\_id="AA010744.1"  
/db\_xref="GI:27361837"

gene  
CDS  
1094..3481  
/locus\_tag="VV12371"  
1094..3481  
/locus\_tag="VV12371"  
/note="COG0072"  
/codon\_start=1  
/transl\_table=11  
/product="Phenylalanyl-tRNA synthetase, beta chain"  
/protein\_id="AA010745.1"  
/db\_xref="GI:27361839"

/translation="MPSSESLREKPNVAVTTIDELTHQITAGLEVDVTLPAQFENG  
VKATVVEGQHPDADKLVTKVDVEEELDITVCAACRGLKVAIVAVLPGD  
FKTKKAKLRGQSHMLCSFTLGIDVEDSDIMELADIPIDMFDFDITANDLVVD  
DLTSNADQFISRGARVEGVNADVPSPVAPVPSIDTVVALEVKAPACPRGLG  
RVKVNVOAKTEPLAQMEKLRGCGISIDPVVDITVFVLLGQGMHAKIDGGLI  
VVRALGQEKITLLDGSBAELNADTLVADHKALAIAGIPGSESGVSEKDVYLE  
CAFPADHIGRRARSYGTLTDSMREPRGVDAVQVAMERDALLVEICGEEVAPV  
AVSEBALPRKNVNAKRTKLDNLGHHIADSDVEIIEELMTVETTLGEGVAVAP  
WRFDAIEQDLVEGRITGYDITENQPAALAKHDOENAPIKRVADLVNDRGTH  
EALTYSFVEPEQQLVPGVDALILENPISAESAVARLQIQGLNTVVMHQRQGP  
VRLFEYGLRPIPCDTNENGRQEPMLAGVIAVIGKTESEHNIDITNVDFDLDGDEAI  
LELSANDKAVSFVAAGPALHPGQSAIIVDEKEIGVIGTVELEKRFKGTATVE  
EIEWSANKRVIPEAVAGLSPANREDIIVVDEAVASGDIYVACIEVGEFLKAAKL  
FDVYVGKVGVEGKKSLAIIALTIOSNRTIEDDINAGVDAIYTHSEKGSALRD"  
complement (3584..4897)  
/locus\_tag="VV12372"  
complement (3584..4897)  
/locus\_tag="VV12372"  
/note="COG1488"  
/codon\_start=1  
/transl\_table=11  
/product="Nucleotide phosphoribosyltransferase"  
/protein\_id="AA010746.1"  
/db\_xref="GI:27361840"  
/translation="MCAPLFOPAIIOGVLDLVYKINMQAAVREYPOQVREYELLVR  
SDNLDLVEEVEEINRLAEIRFDAQALAEKAPVLAEPISYLETFEFPQCOV  
SVGIFRPAQDDCOLRVINGIWEETILFELVMSIISLRNRYMAIPOSQHKYLE  
DKIDPLDSLEKRNITNFRFSEWGTERRRSPAAOKMVLARVLELITGNSHYLA  
QEPNLPITGVVAHERTMHQALVALIHSRVALDKRUEFNGSLGALDTTIGSDIFL  
SDPLDKATVAGVRHDSGSPYWGDKTIAHESLSDIDPTTKTLITFDGIDPRARDI  
CEYFAPRAQISFGITFLANDGMNTNDGTYOPTISWVNAECNCSVAKISDEPE  
KAMCEDIFIFLMNKKORFGLGELVDLKAIFETIKOKRKRIOSVA"  
5176..5889  
/locus\_tag="VV12373"  
5176..5889  
/locus\_tag="VV12373"  
/note="COG1051"  
/codon\_start=1  
/transl\_table=11  
/product="ADP-ribose pyrophosphatase"  
/protein\_id="AA010747.1"  
/db\_xref="GI:27361841"  
/translation="MIVTIDMTCTKIGKGLSVTLKRNPRPQHGMSIRGCFVE  
HDSQGGQPADADPSARRICRQKITHTYPRITSEPNVDGPKRDPKMSVTIAYIA  
LNLQNTIEQIONGCLDQOLAMPALSVQGEVALAFDADLIRLMTAKRAIETYS  
VILFALEKEFLVSDIHAAYAKGVNMTIRRLIDTVIVSANKIASTINKGKGP  
ACVYSLADKHVTFQCLR"  
5902..6546  
/locus\_tag="VV12374"  
5902..6546  
/locus\_tag="VV12374"  
/note="similar to nicotinamide; COG1335"  
/codon\_start=1  
/transl\_table=11  
/product="Amdasee"  
/protein\_id="AA010748.1"  
/db\_xref="GI:27361842"  
/translation="MVDIKKRTASLDVDPQGFSELCENELPVSGALEIVALKONH  
TKAKLTVSRDDEPGLAEAEATPAMKESVGLPEVDIKMNPICVIGTGVBLRPLG  
ALREYFQNMKGIDPDAHYGAFYHQDQLTGTGIEFLRARIDIVAVGLALDPCV  
KKSIEQLVDAEFVILNLSATPAVFPENNQVAVSHLTALGVICIDVDVOLQJHE"  
6962..7249  
/locus\_tag="VV12375"  
6962..7249  
/locus\_tag="VV12375"  
/note="COG0776"  
/codon\_start=1  
/transl\_table=11  
/product="Integration host factor, alpha subunit"  
/protein\_id="AA010749.1"  
/db\_xref="GI:27361843"  
/translation="MALTAKDLAENLFEUGSKDADKTVVFPSEIRKALENEQV  
KISGFNPDLRDKNERPGRNPKTGEDIPITARRVVTTPRGQGLKARVENTLK"



gene  
complement (7347, .8123)  
/locus\_tag="VV12376"  
complement (7347, .8123)  
/locus\_tag="VV12376"  
/note="COG2357"  
/codon\_start=1  
/transl\_table=11  
/product="Conserved hypothetical protein"  
/protein\_id="AAO10750.1"  
/db\_xref="GI:27361844"  
/translacion="MSLFRATLMLVLSRAPAPAIPTAPSNOKORETSONEVS  
KLFKSLGIVGISTOPLOPYDTPVIVSKAHOAFELITVKSTALITDOPPF  
AGVSGQRAKOKIALELINNOBRITDLARATVADVASLISATBALERETRYKVA  
RKSPGSEGRDNLVLVOLPTNLVAEVQLHKAIDVAGSPEDHLYOOLQWEROA  
SETRLNDVELATIKRLSRSHKDIYQAMOPYLITHEAA"  
8413, .9066  
/locus\_tag="VV12377"  
8413, .9066  
/locus\_tag="VV12377"  
/note="COG0500"  
/codon\_start=1  
/transl\_table=11  
/product="Thiopurine methyltransferase"  
/protein\_id="AAO10751.1"  
/db\_xref="GI:27361845"  
/translacion="NRDAPFMNKRAASNOIGFHELVNPLITRMSALAPRETVY  
PLGKTEDLAWLATKHDEBAGELSLVRSFPAHFTPTVTISGHELYOPDELS  
IYAGPFTAPLSKADLIYRALIALPEKREVEYVORIRGLNNGGRLLVSLYPOQ  
ENAGPFSVTOGEIHLFAGMHVTLXODIADHEHPKIAKOGLSRFSBEVYIENDY"  
complement (9119, .10294)  
/locus\_tag="VV12378"  
complement (9119, .10294)  
/locus\_tag="VV12378"  
/note="COG0027"  
/codon\_start=1  
/transl\_table=11  
/product="Phosphoribosylglycinamide formyltransferase 2"  
/protein\_id="AAO10752.1"  
/db\_xref="GI:27361846"  
/translacion="MFGTRBSARVULLSGELKEVAIECORLGEVIAACRYPD  
APAMQVARSYFDMLDASLEBKVIAAPVPEIEALITDLVEBEGALNVVS  
AKATLWNKRSGIRRLAELGLTSPFPAADNYQOFVEAVEASIVCPVAVSSG  
KGQSVTKSPADIEKAMOYAQOEGRTGARVIVEGIDVIDTLLTCVAVDGFCA  
IGHROEDBDYESNQPMQSENAIKAAVTAQVNAVGIGYIGVELFVAGDVIRN  
EVSRRPDGTGLTILSOEMSEPAHRAFTGAPVVKIYQVSPASAVILAGQSENR  
FDGMSDALBOPTOLRFGKPDINGRRLGVTLRBSSTEKAVAAIESAKIKIIV"  
complement (10426, .11334)  
/locus\_tag="VV12379"  
complement (10426, .11334)  
/locus\_tag="VV12379"  
/note="COG0295"  
/codon\_start=1  
/transl\_table=11  
/product="Cytidine deaminase"  
/protein\_id="AAO10753.1"  
/db\_xref="GI:27361847"  
/translacion="MOLQFTKSRFENALASAPESLSRHLPITLADPDASLTVQ  
PEILKOTGMIDNOLVALPFAAYSYAPISFVYGAIVLGLGALYFGAMFEDVY  
OLGOTVHAOSAIISHAMKGGQSLDITINSPCGHQFNNELSSAEIKIQPEER  
EKKHIDVLPDSRGPSDIGIEGALMSOVHGFATEDDDALMORAYENMRSHAPYTHL  
SGVALQTSRGRTIAGYABNAFNSLPPLOVALIQLILAGERFENIQSALVESHKG  
KISHLACTQSTLEALNPDIPVSYLSL"  
complement (11675, .11812)  
/locus\_tag="VV12380"  
complement (11675, .11812)  
CDS  
Query Match 18.1%; Score 467; DB 1; Length 306147;  
Best Local Similarity 59.1%; Pred. No. 3.3e-125;  
Matches 864; Conservative 0; Mismatches 590; Indels 23; Gaps 4;  
Qy 5 TGCACCTTATGACCAATCTTGCACTGCTGACGCGGCGGCACTTGCGCCAGTGGCGG 64  
Db 4407 TCCAGTTTCTCTTAATCTTTGTCAGATTGAGACTGGGGTATTGTCGCCAATTAACGA 4466

Qy 65 CGGTAAGTACTTCAAGATTAAAGCCATGACTAGCGTTTCAATATAAATGCTGTCCG 124  
Db 4467 CGATTGGGAGGCTGTGAGATPAATCACTACATTAAGTTTCAATAGAAATGCTGTGCG 4526  
Qy 125 CAGCTACCTTGAATGGCGATAGCCAGTGGCGTTTGGCCCTTGTGAGG--ATCCCG 181  
Db 4527 CAATGCGATTATGAGTACTGAGCTGCACTGCGCAGTCCCTGTGCGGTGCGAAATACCG 4586  
Qy 182 ATTTCATTTGCGCATCGGTTGAAATAGAAATAGGTTATGATGATGATAAAGTACGA 241  
Db 4587 ACAGAGCTTGTGCTGAGGGTGGAGGCGAAGTTTCTTAATAGAGAGAACTCTGCC 4466  
Qy 242 TTCAATGAGTGCATGCTGCTCTAATAATAACATGTGGCATCCGAAAAGCGAATGAA 301  
Db 4647 GTTAAATAGGGCCCTTCTCTGCAAGTAAAGCCAACTGGGCAACATCAAGCAAGCTCT 4706  
Qy 302 GCCAATGATGATTTCTTGGCGTACTTCTCTTAATTAATGCTAATGCTTATGATCACTG 361  
Db 4707 GCGAGGCGTGAATTTCTCTCAAGAACCTTCTCACTTAATCAGACAAATTTATCAAGAG 4766  
Qy 362 CGCAATCAATTCATAGCGCACTCAATCCGATACAAAGCAATGAACGGCGTGCATC 421  
Db 4767 CGAAGCATGTTGTTGAGGCACTTGAGTTTGAAGTAAATGATGATACGGCGCTTGATC 4826  
Qy 422 ATATTGATTTTATAGGATCAAGATCAATTAATCTGCGATTAAGAGAGAAATAGG 481  
Db 4827 ATATTGATTTTATACATCTAATTAATGAAGAACCTTGAAATATGCTGTTGAAATAGT 4886  
Qy 482 CGATCGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 541  
Db 4887 GCGGCGGACATGCTGCTGTTTACCCCTG-----TTAGTTTAAAGAAC 4929  
Qy 542 AACGTTGCTTAATCTTGGCGCAATTAAGTACCTTGAATGTTTCACTTTTGTG 601  
Db 4930 ATTTTGTCTTATATGTCG-GAACCTTAATGAGAGATGATGATGATGATGATGATGATG 4988  
Qy 602 ACAACCTAGTCACTGTTATTTGGCTTATTAATTA--TGGATATGAGGGGTAAAG 659  
Db 4989 ACATACATCAATGCTCTGCGAGGAGATGCGCGGCTAATTGATTTCTTAATTTGAAG 5048  
Qy 660 ACGTAGTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 719  
Db 5049 GATGAGCGGCGCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5108  
Qy 720 TTATTTACATTTGACCATCATCATGACTTACCTAAATTAAGCCCGTGTATTAAGGGA 779  
Db 5109 CTTTCCATTTGACCTTAATGAGATGCTCTCAATTTGATGTCAGATTTATTTAAGAG 5168  
Qy 780 AGCCATTATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 839  
Db 5169 TTGCGGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5228  
Qy 840 GCTTTATCGTGAACGCTCTTAATCAATTCGCGCAATGCGGCAATGCGGCAATGCGG 899  
Db 5229 AGTCTGCTGATTAAGGCAACAAATCTGAGAACCAACAGATGAGATGAGGCTGATTC 5288  
Qy 900 TGGCGGATAGTATGACGAGATATGACCGCTCATGTTGAGAACTGTGATGATGAGGA 959  
Db 5289 GGGAGGATTTGTGTTGAACAGATTTATCCAAAGGGAGGACCAACCGCTGACCGTGA 5348  
Qy 960 TTTTGTATGACGAGACGAGATTTTGTGGGCAAAAGTCCATCTTATCTTAATTTTAT 1019  
Db 5349 TTTTGTATGACGAGACGAGATTTTGTGGGCAAAAGTCCATCTTATCTTAATTTTAT 5408  
Qy 1020 CAGCGATCCGCTGTTGATGAGCAACCCAAACCGCATCCGATGTTGAGTGTGATGAT 1079  
Db 5409 CAGTGAAGCGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5468  
Qy 1080 TTCCCATTAAGCTTTATTAACCGTGAATGTGAACCAATGAAGATTTGATGATGAT 1139  
Db 5469 CGCCCATTAAGCTTTATTAACCAATGAATGAACCAATGAATGAATGAATGAATGAATGA 5528

Qy	1140	CCCCAGGCGCGTAAATGTTGATCTTCATCTTACTTACCAAGAAAGAAAGCCGCTGAC	1139
Db	5529	TGATCAACAATTTGGCTTGCTTCGCCCTTGACGCGTGCTGCAGAGGGAAGTGCCTTGGC	5568
Qy	1200	TTTGTGATCATGTCCGCCAAATTACAGATGCGTGGCAAAATTACGCGCTGGCGTTGAATA	1255
Db	5589	CTTTGACCCATCCCGATTTTAATTCGCTCGATGAGCAACAAGTTCGCTGCGCGGATCGAATA	5648
Qy	1260	CACATCCGTGTACTATTTTCATTAGAAAAGAGTTTATAGTGGCGGATATATATGATGC	1319
Db	5649	CACATCGGTGGTCTTTGTGACCTCGAAAAAGATTTCTCGTTTGTGATCATCATTCACGC	5708
Qy	1320	CTAGCCCAATTGGCGCGTCGAAAGTTAAATGCGATGACCTATTAAGCCGCGCTGATCAATAC	1379
Db	5709	CTAGCCCAAGTTGGTGTGATGTGAACCGCATGTGCGATCAAGCACACTTTGATGACAC	5768
Qy	1380	CGGGGTGATCGTCACTATACCAATTAATAATGACCGCATCTTGTAAAGGCAAAAGGCCAAAC	1439
Db	5769	CGGTGTGATCGTAGAGCGGAATTAATAATGTCTCGACCAACAAGGGAAGGAGGCAAAAC	5828
Qy	1440	AGCCACCGTTTATCGTCTTGCAAGTCATGAAGTGCACATTTTCAACCTGTTAAG	1496
Db	5829	CGCTCAGTCTAATAGCTTAAGCGGATTAACAAGTAACCAATTTTCAAACTTGCTTGGC	5885

gene  
CDS  
gene  
CDS

AVHGACVKKGEVRSFPEVDVGRHVAVDTLLGEMMLINQESGADKI FYTTGRLTSMVTK  
VAMGCIPIILSRSGVTOMGDLIAQPGITTTIARAKGLRFQVFTGGEIITIDVQGES"  
1151. .2908  
/gene="VV1707"  
1151. .2908  
/gene="VV1707"  
/codon\_start=1  
/transl\_table=1  
/product="methyl-accepting chemotaxis protein"  
/protein\_id="BAC94471.1"  
/db\_xref="GI:37198637"  
/translation="MSGSPVQKRPFFSFRKULLINVTLLGAAVAIYENNSAGH  
LSLEHASTELSSIDLIMRHEKHYLIARSRKVLSPATVSGOLQIDHSDVIA  
ASESTALDDQACQIVTLLSQYQTFHQAQNLIDALQESTLPRABRMKTLVTSND  
GELTAQFDQLDHDLTFTVEPTQEHKVALLEGSLQPAFEADLSPATDTSNTLSRTY  
SAITLGTANEGEALRSVAKHTBEAIDTDLTIQSGVEQTADIRSHLQMGVDFE  
VAILLSLIYVIGRSILSRIRKALINDMIDANGSDLTIVRNANKSDEIAQISRFDFE  
TSHADNINKOLASVMDVIGSESSCSBOAQRSMQARVNSEADKALTATGKRIID  
TLAISIESQAOIALAEGSRETIAVSAIQGLAEQNILALNAALIAAAGSEGRGE  
AVAADEVRQSLRLINDSTLQIESITVQALTKTIACTVAKMADSVQCAATTVDPHVE  
AIEGISTQITKEMFPMNTQIATASEESQMSAEIDRNTIQALQGLDGTIEIVSGVRCS  
EÖVDSVSHKEKIVAQCFY"  
complement (3018. .4610)  
/gene="VV1708"  
complement (3018. .4610)

RESULT 3	AP005337/c	241900 bp	DNA	linear	BCT 04-DEC-2003
LOCUS	Vibrio vulnificus Y016	DNA, chromosome I, complete genome, section 8/4.			
DEFINITION	AP005337				
ACCESSION	AP005337	BA000037			
VERSION	AP005337.1	GI:37198635			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	Chen,C.Y., Wu,K.M., Chang,Y.C., Chang,C.H., Tsai,H.C., Liao,T.L., Liu,Y.M., Chen,H.J., Shen,A.B., Li,J.C., Su,T.L., Shao,C.P., Lee,C.T., Hor,L.I. and Tsai,S.F.				
TITLE	Comparative Genome Analysis of <i>Vibrio vulnificus</i> , a Marine Pathogen				
JOURNAL	Genome Res. 13, 2577-2587 (2003)				
REFERENCE	2 (bases 1 to 241900)				
AUTHORS	Chen,C.Y., Wu,K.M. and Tsai,S.F.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-MAY-2002) Shih Feng Tsai, National Health Research				

/gene="VV1708"  
 /note="containing duplicated ATPase domains"  
 /codon\_start=1  
 /transl\_table=1  
 /product="ABC-type transport system, ATPase component"  
 /protein\_id="BAC94472.1"  
 /db\_xref="GI:37198638"  
 /translation="MISTANTVQFGAKPFLNYSVKGSGNRYGLIGANGCKSTPM  
 KILSGEELPESIGANSYDENERVAKLNDQDFYEFTVIDVYIMKHKLAMVYKQDRKDI  
 YSLSEMSIEDPAGVADLVEVEPABNDGMAEKACQELLIAGVIMPEOBYGMSVAPVPM  
 KLRYLALQLEPADPIMLDEPTNNLMDLRTMTLEETLNNACMTIM1ISDRHPLNSV  
 CTMHADQGLGLRYPGNYDEWPAASQARRELLISDNAKKCAQIAEIQTVVAFASNA  
 SKAAQSTRAQIDIKILDEYKASSRNPFRFESQELFNALIVNLISQGFEDLFF  
 SDPNAIPVEYGRVAILGNGGVKTTLLNTAGVTEPTEGKSENNATIVYQDDHAK  
 EPABDNLMMNGMGQRGGDEBQYVRGFLGMLGQDDIKSVYVLSGGGQGRVLLH  
 LMEHPNNLNLDEPTNMDMESISLNNALNLYQKGLTFVSHDRVFDLSLNRITLIR  
 DGKLTDFPGTSEYLSKRGVGS"  
 4834..7179  
 /gene="VV1709"  
 4834..7179  
 /gene="VV1709"  
 /note="identified by Genemark and Glimmer2"  
 /codon\_start=1  
 /transl\_table=1  
 /product="hypothetical protein"  
 /protein\_id="BAC94473.1"  
 /db\_xref="GI:37198639"  
 /translation="MOKLRLRFVCFTLVFPAPATPAQMOYAAQPAQVSAQALSIESD  
 SALDPEPMGEKHEIKYNOLLASVLELQTEGDAFKAALQYKAPAPQWLIVENQYQ  
 LTNISGSSKRLILQNDTASRBQLTGEGPQVQPEALATLNTLNVLPQRFSE  
 BALIKNTISVPIVWTGSKRPAYTYLLMMMAANSKRITLFRPSOLENTAPELWPK  
 FTWISRNKRIAMVILMTSLRIIASIPSIQHVYIIEFTFWILGSGIASISFTLEA  
 LVYSSSSKDDVALRLATLRTIRYVSAIVAGVITLISRLIKGTYITWISLHFKVY  
 LNTVSSSLDKMKVKEEVVDRIITERPVWQVAKIKQDSFLLIPATVIAAWLISHKRV  
 HRIATLSNYPFSCALAYLRIEVAQKTSRNDGSDSLVIRIGDAQOYLYLPQODEE  
 MIDVADSIQMLSRVLLDTPSPALCVLSGERGVGTSSYLKILGKRNKATPLVLYNC  
 GYPLRLKQALGLGLAESSTEVQILAFRKSETPYLLAVDAQRLVPMVQGLGLMKM  
 LTNLIRKQKONHRYVMAIERASMRVFNARGEERLIPWMSLIPMNKQISALDSRKY  
 NQDLENPLSPGIVYPMQODEETSEERALACQGYRTILMHYSDNPVALRPFSLIR  
 NKRSDDVAVRLFHAPSKELBQMPKRYLAVLSIVQLEVASPELSBCYQLSLAEVYI  
 GTRFESRSGEIANNEDEKARVSDHMFRIITVLDROHLVA"  
 7181..8062  
 /gene="VV1710"  
 7181..8062  
 /gene="VV1710"  
 /codon\_start=1  
 /transl\_table=1  
 /product="small-conductance mechanosensitive channel"



QY	900	TGCGGGATAGTGTATGACGAGATATGACCGCTCATGGTGGAGAACCTGTGATGAGA	959	TITLE	Direct Submission
QY	201543	GGGAGGATTTGTGTTTGAACACATTTATCCCAAGGAGAGACACCGCTGACGCTGA	201484	JOURNAL	Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp, Fax:81-6-6879-2047)
QY	960	TTTTGATGACGAGACGAGATTTTGTGGGCAAAAAGTCATTAATCTTAATTTAT	1019	COMMENT	This clone was isolated from a patient presenting with acute gastroenteritis.
QY	201483	TTTTGATTCGCGCGAGCGTCTGTCTGTGGGCAAAAATTCATTAATCTTAATCT	201424	FEATURES	Location/Qualifiers
QY	1020	CAGCGATCGCTGTGGTATGGACCAACCCCAACCGCATCCGAATGGTGGATGTCAAT	1079	source	1. 303450
QY	201423	CAGTGAAGCCATGTGTGATGGTATGTAACCAACCGCATCTCTATGTTGAGCGTGA	201364	gene	/organism="Vibrio parahaemolyticus"
QY	1080	TTGCCATTAAGCTTTATTAACCGGTGAATGCAAAATAGAAATTTGTATGCA	1139	CDS	/mol_type="genomic DNA"
QY	201363	CGCCCACTATGCTCTGTTATCAACCATATGCAAAATTCAAACTGTGGTTGG	201304	gene	/strain="O3:K6"
QY	1140	CCCCGAGCGGCTAATTTGTTGATCTTCACTTTACTCAAGAAAGAAATCCGCTGGC	1199	CDS	/sub_strain="RIMD 2210633"
QY	201303	TGATCAGCAATTTGGCTTGGTTCGCGCTTGATTCGGTCTGCAAGGAGAGTGGCGCTGGC	201244	gene	/db_xref="taxon:670"
QY	1200	TTTTGATCAGTGGCGCAATTCAGCATGGTGGCAAAAATTAACGGCTGGTGAANA	1259	CDS	/chromosome="2"
QY	201243	CTTTGACATGCGCATTTAATTTGCTCGATGACCAAAATTCGCGCATGCAATA	201184	gene	81. 1574
QY	1260	CACATCGGTGACTATTTTCATTAGAAAAAGATTTTATGCGGATATTTATGATGC	1319	CDS	81. 1574
QY	201183	CACATCGGTGCTGTGTTTCACTCGAAAAAGATTTCTGTTTCACTAATTCAGCC	201124	gene	/gene="VPA0300"
QY	1320	CTACGCCAATTTGGCGCTGCAAGTTAATGCGATGACCATTAACCGCGCTTGATCAT	1379	CDS	/gene="VPA0300"
QY	201123	CTACGCCAATTTGGTGTGATGTGAACCGCATGACATCAACCGCGTTTATGACAC	201064	gene	81. 1574
QY	1380	CGGGGTGATGCTGATACCAATAAATGCGCGCATCTGTGAAGCAAGACCAACC	1439	CDS	81. 1574
QY	201063	TGCTGTGATGTGACCGCAATTAATTCCTGTCGACCAACAAAGGAGAGGCAACC	201004	gene	/complement (1835. .2314)
QY	1440	AGCCACCGTTATCGTCTTGGCCAGTCAAGAGTCACTTATTTCAACTCTTTACG	1496	CDS	/complement (1835. .2314)
QY	201003	CGCTCAGGTCTATAGCTTACCGATTAACACGTAACCTTCAACTTGGCTTGG	200947	gene	/complement (1835. .2314)
RESULT 4	AP005085/c	303450 bp	DNA	linear	BCT 05-MAR-2003
LOCUS	AP005085	303450 bp	DNA	linear	BCT 05-MAR-2003
DEFINITION	Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 2/6.				
ACCESSION	AP005085	BA000032			
VERSION	AP005085.1	GI:28808465			
KEYWORDS					
SOURCE	Vibrio parahaemolyticus				
ORGANISM	Vibrio parahaemolyticus				
REFERENCE	1	Nasu, H., Iida, T., Sugahara, T., Yamachi, Y., Park, K.S., Yokoyama, K., Makino, K., Shinagawa, H. and Honda, T.			
TITLE	A filamentous phage associated with recent pandemic Vibrio parahaemolyticus O3:K6 strains				
JOURNAL	J. Clin. Microbiol. 38 (6), 2156-2161 (2000)				
MEDLINE	20295086				
PUBMED	10834969				
REFERENCE	2	Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Nishijima, M., Nakano, M., Yamashita, A., Kudo, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H., Hattori, M. and Iida, T.			
TITLE	Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V. cholerae				
JOURNAL	Lancet 361 (9359), 743-749 (2003)				
MEDLINE	22508454				
PUBMED	12620739				
REFERENCE	3 (bases 1 to 303450)				
AUTHORS	Oshima, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T., Honda, T., Shinagawa, H., Hattori, M. and Iida, T.				



```

QY 663 TAGCATACCAACATTAAGTACTCTTGTATCTAGTATGTTGTCACAAAGTCTTA 722
Db 8096 TGAAGATTATAGGATCTGTATGCTCAATCGATTTCTGTTACAGAGGTGACAAATTTCT 8037
QY 723 TTACATTTAGACATCATCAGCACTTACCTTAAATTAAGCCGTGTTTA----- 772
Db 8036 AACCACTGCTTACTTACCACTGATTTGACCACTTGTATGCGCTGATTTACTTAAATC 7917
QY 773 ---TAGGAAAGCCATNTAGATTGTCACTATGCAATGATATGTTGTGCTGCT----- 821
Db 7976 AATATTAGAACCCACCATGATCTGATCTGATGCAATTTCCATTCAGGCTATCTGCT 7917
QY 822 -TGCGCCGAATCTATCCAGGTTTACTGTTGAAAGCGCTATATCAATCGCCAGATT 880
Db 7916 GTGCGGATTAAGGATTTGAAGTGTGCTGATTAAGTTCAATCCGAATGGCTTACC 7857
QY 881 GTGCTAATAGGCAATTTGCTGCGGGATAGTGTATGACGAAGATATGACCGCTATGCTG 940
Db 7856 ATGGGTGCTGGGCAATTAACGGGCGGGTGTCTTTGATTAAGACTTGAACGAGAGCG 7797
QY 941 GGAACCTGTCATGATGAGATTTTGTATGACGAGACGATATTGTCGGCAAAAAGTCC 1000
Db 7796 GACGTCCGCGGATTAAGAACTTTGAAGCGCGCGCGCGCATTTGCGCGAGAAAGATTC 7737
QY 1001 ATACTTATCTATTTATTCAGCGATCCGCTGTTGATGCAACCCCAACCGCATCCGA 1060
Db 7736 ATACCTATCTCTGATCACTTCAAGGAGGCTTATGACGAGAGCCCTTAAGCGCATCCGG 7677
QY 1061 ATGGTTGAGTGTCACTATTTCCCATTTAGCGTTTATTAACCCGGAATGTCAAACAA 1120
Db 7676 AAGATTGAGCTTAAACATCACCAATAGCATTTAGTGAACGCAACAGTCAGCAAA 7617
QY 1121 TAGAAGATTTTGTATCGACCCCGAGCGCGCTTATGTTGATCTTCACTTACTCA 1180
Db 7616 TCAACATGCTGCGCGCGCGCGAGTGCAGCTCAATAGTCCCATTTGACGCGATCTCTCA 7557
QY 1181 AAGAAAGAAATGCCGCTGCTTTGTATGATGTCAGCGCAATTCAGATCCGTCGCAAAAT 1240
Db 7556 ATGGAGAAAGAAACCTTAGATTTGACCAACAAAAACATCGAAAGCATGGCGAAAC 7497
QY 1241 TACGCGCTGCGGTTGAATACATCCGCTGATCTATTTTCACTTGAAGAAAGTTTATG 1300
Db 7496 TCGGCTGCTTCCATGATGATCAATCAGTGTCTGCTTGGCACTGATTAAGATTTTGG 7437
QY 1301 TGGCGGATTTATTTGATGCTTACGCCAAATTTGGCGTGAAGTATATGATGACATTA 1360
Db 7436 TCGCAGACATCATTTCTGATATCAAGAAATTTGGCATGACATCGACGCGCATGACATTA 7377
QY 1361 AACGCGCTGATCAATACCGGGGTGATGCTGATGACATTAATTAATGCGCGCATCTGTA 1420
Db 7376 ACGCGCGTTTATGATCACTGAGTGTGTTGAAGCCGACCAACAAAGTCCGCTCCACAAAC 7317
QY 1421 AAGGCAAGAGCCAAACCAAGCCACCGCTTATGCTTTGCGAGTCAAGTCAAGTCTTAT 1480
Db 7316 AAGGCAAGAGCCGATTAAGCCAGCCAGTGTGTACACCGCTCAAGGATGAAGTGAATCTT 7257
QY 1481 TTCAAAACCTGTTAGAGGTTTACTGTTGAAA 1513
Db 7256 TTCAAAACCTGTTTGGTGGCTTAAACGCAACGA 7224

```

```

RESULT 5
AP005349 247450 bp DNA linear BCT 04-DEC-2003
LOCUS Vibrio vulnificus YJ016 DNA, chromosome II, complete genome,
DEFINITION
ACCESSION AP005349 BAO00038
VERSION AP005349.1 GI:37201329
KEYWORDS
SOURCE Vibrio vulnificus YJ016
ORGANISM Vibrio vulnificus YJ016
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.

```

```

REFERENCE
AUTHORS
1
Chen, C.Y., Wu, K.M., Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L.,
Lin, Y.M., Chen, H.J., Shen, A.B., Li, J.C., Su, T.L., Shao, C.P.,
Lee, C.T., Hor, L.I. and Tsai, S.F.
TITLE
Comparative Genome Analysis of Vibrio vulnificus, a Marine Pathogen
JOURNAL
Genome Res. 13, 2577-2587 (2003)
REFERENCE
2 (bases 1 to 247450)
AUTHORS
Chen, C.Y., Wu, K.M. and Tsai, S.F.
TITLE
Direct Submission
JOURNAL
Submitted (29-MAY-2002) Shih Feng Tsai, National Health Research
Institutes, Division of Molecular and Genomic Medicine, 128,
Yen-Chin-Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China
(Yen-mail:petal@nhri.org.tw, Tel:866-2-8146-1041,
Fax:866-2-2789-0484)
COMMENT
This sequence was determined by the Sequencing Core of the National
Yang-Ming University Genome Research Center (YKGC;
http://genome.ym.edu.tw).
FEATURES
source
1. 247450
/organism="Vibrio vulnificus YJ016"
/mol_type="genomic DNA"
/strain="YJ016"
/db_xref="taxon:196600"
/chromosome="II"
/complement(73..1092)
/gene="VVA1126"
/complement(73..1092)
/gene="VVA1126"
/codon_start=1
/codon_table=11
/product="outer membrane protein N, non-specific porin"
/protein_id="BAC97152.1"
/translacion="MKKILALVAATPQTQAVVEVYNDGTFISIGHVSVALLED
SKOGDVGTAAPRINFATODLNGFTADAGSINVLADGEMALSTRLVIGLTH
BSLGRVGTQWAPRYSVAGVADMPLEAFNDPIYDGHGLDGEAEVSVYAAEFLG
EAGKIALGCMGRKTDNNNSGNRAQIALNSINLVNVAVNTGDDVYVVGKTI
SESHVLSATYSGKGLYIAGYANMEYNSGDGULIKDTSAYEPLAATLWNGINLS
VNTESVIDELSLRVYSIAVDALNIEYFKFQFAGYQPDLSGSDYKEADKWLGLG
ARYLL"
/complement(1331..1519)
/gene="VVA1127"
/complement(1331..1519)
/gene="VVA1127"
/note="identified by Glimmer2"
/codon_start=1
/codon_table=11
/product="hypothetical protein"
/protein_id="BAC97153.1"
/db_xref="GI:37201331"
/translacion="MKKILALVAATPQTQAVVEVYNDGTFISIGHVSVALLED
SKOGDVGTAAPRINFATODLNGFTADAGSINVLADGEMALSTRLVIGLTH
BSLGRVGTQWAPRYSVAGVADMPLEAFNDPIYDGHGLDGEAEVSVYAAEFLG
EAGKIALGCMGRKTDNNNSGNRAQIALNSINLVNVAVNTGDDVYVVGKTI
SESHVLSATYSGKGLYIAGYANMEYNSGDGULIKDTSAYEPLAATLWNGINLS
VNTESVIDELSLRVYSIAVDALNIEYFKFQFAGYQPDLSGSDYKEADKWLGLG
ARYLL"
/complement(1559..2050)
/gene="VVA1128"
/1559..2050
/gene="VVA1128"
/codon_start=1
/codon_table=11
/product="thiol-disulfide isomerase and thioredoxins"
/protein_id="BAC97154.1"
/translacion="MKKILALVAATPQTQAVVEVYNDGTFISIGHVSVALLED
SKOGDVGTAAPRINFATODLNGFTADAGSINVLADGEMALSTRLVIGLTH
BSLGRVGTQWAPRYSVAGVADMPLEAFNDPIYDGHGLDGEAEVSVYAAEFLG
EAGKIALGCMGRKTDNNNSGNRAQIALNSINLVNVAVNTGDDVYVVGKTI
SESHVLSATYSGKGLYIAGYANMEYNSGDGULIKDTSAYEPLAATLWNGINLS
VNTESVIDELSLRVYSIAVDALNIEYFKFQFAGYQPDLSGSDYKEADKWLGLG
ARYLL"
/2203..3180
/gene="VVA1129"
/2203..3180
/gene="VVA1129"
/codon_start=1
/codon_table=11
/product="putative outer membrane protein"
/protein_id="BAC97155.1"

```







QY 1915 TTTCGGTTAAAAAGACAGACGAGCAATCCGCTATCCACAAAAATCAGCAGTATTGCA 1974  
DB 146083 CATGGGCTCAATATGCAATGAGATGTTTGGATTTCCTCAAAAGCTTGGAGATTTGCA 146142  
QY 1975 GGAACGATACCTTTAGGGGCTGGCGACCTTTTACTGCGCGATCCACTG--GACTCCGAT 2031  
DB 146143 GGCACGATTCCTAGTGGGTGACAGACCTTCTTACTGATGATGATACACAGCAATGCTGAT 146202  
QY 2032 GGTACCGTACAGTGAAGAAACCAAAATAGCTGGCATGACAGATCATATGCGATATCC 2091  
DB 146203 GGGAGCGTACCGGTGAGAGACAAAGATCTAGGTATGACTGACCAATGTTTGAAGAAA 146262  
QY 2092 ACCACTTCATACGAGATGCTGTTTATCATTCCTGTCGCGACCAATGACCACTTCT 2151  
DB 146263 AATA-GCGATACAGCATGATATATATAGCCGATATTCAGATCAGATCAGATCATTTGAT 146321  
QY 2152 TCGTATATGACCGCTTCCGCG 2171  
DB 146322 CCAGCATTCCTCACTTCAGGC 146341

RESULT 6  
AE016810 302331 bp DNA linear BCT 29-SEP-2003  
LOCUS  
DEFINITION Vibrio vulnificus CMCP6 chromosome II section 3 of 6 of the  
ACCESSION  
VERSION AE016810 AE016796  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
14500463

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
14500463

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
14500463

FEATURES  
source

gene  
CDS

1.302331  
/organism="Vibrio vulnificus CMCP6"  
/mol\_type="genomic DNA"  
/strain="CMCP6"  
/db\_xref="taxon:216895"  
/chromosome="II"  
complement(70..1368)  
/locus\_tag="VV20549"  
complement(70..1368)  
/locus\_tag="VV20549"  
/note="COG325"  
/codon\_start=1  
/transl\_table=11  
/product="Chitinase"  
/protein\_id="AA007498.1"  
/db\_xref="GI:27358549"  
/translation="MKKTLITAAALPSAFAGSVAADQKVVAGVAPADMOYANANPYTK  
DPAKLTIVITVIAPIAMCGPHGASSETVOKIVAKCGKEPESALIVDTALAKDNG  
KIVKVPFKGHAQLAELKQRPDIKIIPSGRTMSPPHAKADPAMAKQPSKIV  
ELIKQYDFPDGIDLMETIPGGGLITSPNPAITKIDQKAERDAFTYLVKTLRADL

gene  
CDS

DALEKTNREVELSTAVGVAKAAQIDMKAAQPYLTNNPAMTYDELAGGCGOQTGHTTN  
LHATERSWNGADIVP INQNTIELGIPSEKLVGAAPYGGMGQGTDDYDQAKGLIS  
EGAGRGNGENGVPMWMDLVNRYGKQGVVAVDQSOAPVYAMNPKYKFIIFEDQRS  
IRAKAKAQSLGGLFTYELSGDPSGKLVDMALMTSK"  
complement(1510..2502)  
/locus\_tag="VV20550"  
complement(1510..2502)  
/locus\_tag="VV20550"  
/note="contains sigma factor-related N-terminal domain;  
COG2390"

gene  
CDS

2966..3916  
/locus\_tag="VV20552"  
2966..3916  
/locus\_tag="VV20552"  
/note="COG0176"  
/codon\_start=1  
/transl\_table=11  
/product="Transaldolase"  
/protein\_id="AA007500.1"  
/db\_xref="GI:27358551"  
/translation="MSNKLQRLKLTVVADTGEIDAKIKYQPEDATTPSLIKRAQ  
IAEAPLIDASIEVATQSDDKAQIQIDCNLAVNIGKILKTIKPISTVDALIS  
YMEGSAVARQIVKXNDAGITNDRILIKASITWEGIPAAIIEKEGNCMLTLFS  
FQARACAGVGLISPEVGRIMDYKAGEIDPEAOEDPGLSVYKINYYKEGYK  
TYMGASFRNIGLIEIACGDRLLTAPALAEBAEGVEKLVDSKGAARBPAMT  
HAEFLMEHNDPMVAEKLAEGRINFAVDGKLEAMIAAL"

gene  
CDS

3998..5989  
/locus\_tag="VV20553"  
3998..5989  
/locus\_tag="VV20553"  
/note="COG0021"  
/codon\_start=1  
/transl\_table=11  
/product="Transketolase"  
/protein\_id="AA007501.1"  
/db\_xref="GI:27358552"  
/translation="MDRQLANAIAPALSDMGVQKANSGRHAPMGVADIAEVLWNGH  
NHPSPNEMADRDRLPVLNSHGSMILYSILHSGEISIDDLKNEFOLSKTPGHEEY  
GVAPGIEITTPGLGGITVAGVAAAEKLAQFNGEISIDDLKNEFOLSKTPGHEEY  
ISREACSLAGTIGKTLAPMDNGISIDGHEGMSDDTPRFRAGVHVI PAYDGH  
NABATINAIEMAKADRPILCTKTITIGRSPKSGSHDCHAPLGAETIATRXEIG  
MEHPPELIEQVEYAEWSAKAEAAANNEKFAIEKATPELAERFRVNGELPRAO  
MEKKAQIILADQANPANTASRKASQNALAEAGKLPBWGSGADIALPNTLWSSK  
SLEASDFSGNVIHYGVREFGNTALINGIALHGGFVYGATPLFMFEYANANRMAALM  
KVONIOVYTHDSIGLEDGPTQPVQVQIASLTLTPMSYWRCDQVESAVAKLAIER  
KQGPSALIFSRONLQOGRASQVADIAAGVILKMSDGKPELILATGSEVELAVKA  
ABQLTAEKKKRVSMPTDTPDKODAAVRAVLPBVDVRAIARLEAGIDPMYKYVGF  
DGRITGNTTGESAPADQLFEMGFVENVVTAKEELLA"

gene  
CDS

6301..6864  
/locus\_tag="VV20555"  
6301..6864  
/locus\_tag="VV20555"  
/codon\_start=1  
/transl\_table=11  
/product="Conserved hypothetical protein"  
/protein\_id="AA007502.1"  
/db\_xref="GI:27358553"  
/translation="MPGAEDLTVRVFTQCGPLENAVVFLEKDALLOGLQPMVAEL  
SCHRRFTPRNVAVITKGSVAEPNDRVRRHHVYSFNATKPELIKYSQSPKRPVVDK  
AGVVEICGNIHSMAMLIIVSDPLPATDNGCEVAFITQOQADTYQIEVHSHALPGS  
PIYTGSTRILTANNAAVEIIMPALGSL"  
6861..9254

CDS  
/locus\_tag="VW20556"  
6861..9254  
/locus\_tag="VW20556"  
/note="Other COG candidate exists; contains a membrane domain, an EAL and a GGDEF domain; COG5001"  
/codon\_start=1  
/transl\_table=11  
/product="Predicted signal transduction protein"  
/protein\_id="AAO07503.1"  
/db\_xref="GI:27358554"  
/translation="MNSLFGHYLRISRRPTFLALISIQVSYALVQVYDAHI  
REIRLEAVSRVWQDLIQASFRLOQNALDLADFEPRVAVADQDTTSLQNN  
AARVADSIADIDHQMALQALSEDDISESRKELVTAIKQSTAPYHLMYKQPTQ  
FYLAIPARIRIGVYLFGEIETHORISLASQAGVDAALISLSISNIIYVSEAOI  
KAYFSIAERIPASTMODRAIDRVVTEETELCPNQHIVYVPELSDASLAR  
DKLASOYIMVSGIAIVLFTALALSHRIVPLITLTTPKATREDAQVATSECTGP  
DEIGQITRGFDQMRERLOKQKTTIQLAIFDTLTLPREHFRQKQELBIKSI  
SVITINDRFKINDVIGHGDVAVLQMAQRLVTSILTFAPARISGDFVEIKT  
GSDLEADSVTRIKORLDIPEVNDPDIISVIGVATPSSGDGDVSLNLSQIAMS  
ARKKEQVFPYDSALVTAIPENTISLSIRPALQNELVYLQPKVATTVISSAER  
LIRMOHPERGLAPPEFVPAQOTGYVELTRMMIQOYIOHRELQPLSGQLTSVNL  
STWDLINPOLDLIDHMLSEYHPVPSAGCLEITTSALIMEDETSLETITLISOMFHL  
STDDRCYSSISGVQRLPYNELKIDKSPVAMLENPNDKIVSTTHIASLQVY  
AEGVSYDLFTSRLELGDRIQICVSKPLADEPTLAWEEMOSNPYVENWKQG"  
complement (9304..10866)  
/locus\_tag="VW20557"  
complement (9304..10866)  
/locus\_tag="VW20557"  
/note="Other COG candidate exists; COG0840"  
/codon\_start=1  
/transl\_table=11  
/product="Methyl-accepting chemotaxis protein"  
/protein\_id="AAO07504.1"  
/db\_xref="GI:27358555"  
/translation="MSALISNSAHEITLKEHOLVSTTDLKGVITTCNEAFRAEBS  
BOELIGENHIVIRHSDMPKRAFGDMQMLKQCKMRGIVKRTSGGAVDAVTPSI  
YENGWMSGYQSVKPKREKWEYIASKAHSILDERNRKQSLKNSLRATILGSI  
SAPVLAQMLSDGTSTLANWPLTTLALFRQELIDTPQQLKALQSEPSVSLIS  
KGNQSIADFHIKASARIKVLGEMTSARPLQQLADLEATVEESQASQASIS  
RQVQQTDSMEVAAQSVSSIQDADRLLEETORCVAKQSIIDTHRELMQIAEK  
ATATYQLSBOAKQSVQMEIEGIAEOTNLALNAAIEARAGQSGGVVADVER  
ALSGRTNATLIRSIDTMLATIQSGMDLLENKQCTESSEVAIISAEISVEBM  
MRSQOLATYIASADNOMQLSSQVNHHSIVTTSQNLAATYSVENSGKLSRPD  
DPYQLAYRFEK"  
complement (11177..12229)  
/locus\_tag="VW20558"  
complement (11177..12229)  
/locus\_tag="VW20558"  
/note="DAH; COG0722"  
/codon\_start=1  
/transl\_table=11  
/product="3-deoxy-D-arabino-heptulosonate 7-phosphate synthase"  
/protein\_id="AAO07505.1"  
/db\_xref="GI:27358556"  
/translation="MRTQDDVRIINNVKELLPVAVLEKYPATEASTTPOSRNDIHN  
ILIGDRLIVYGCSDHDPRAIYERKRLADELKGELIEMVYEFKPTTNG  
WKLINDPYLNDYTKLNDGLRNGRLDLIDWGLPTASEBLITTPYVADLSKGA  
IGARTESQVHEHLASGLSCVGFNGIDGNIKIASDILRASSHHLSSTRKQYRA  
IVETAGNDPCHILIRGKEKSPSAHVAVSEBELTSLPKVWIDSHANSKQYR  
QMLVADVSEBLAGEBAIFGVMIHSHLVESRQDLDGKAATYQCSITDICTGMEDE  
TVLRQLANAVABRAK"  
12652..13725  
/locus\_tag="VW20559"  
12652..13725  
/locus\_tag="VW20559"  
/note="peptidoglycan-associated (lipo) proteins; COG2885"  
/codon\_start=1  
/transl\_table=11  
/product="Outer membrane protein"  
/protein\_id="AAO07506.1"  
/db\_xref="GI:27358557"  
/translation="MKKLAVTLISAVLASSGFPSPATANAAMSDLYFGKIGYVNLDDAC

Query Match  
Best Local Similarity 62.9%; Pred. No. 1.8e-55;  
Matches 390; Conservative 0; Mismatches 226; Indels 4; Gaps 2;  
8.9%; Score 230.4; DB 1; Length 302331;  
YVSPDCDDSDAGFLYYGVNFSKHISAEIGIDILGHEHNSNGSESAKLAAYTLAP  
KPSLPIANEKIDAPAKIGAAIMTYIGDDKDLVPTSLGIEYALTSFPARAEYQYQDMS  
DDIVKMDANFPGIGITLYLFGASSTATYAAANNAIVEPRAEYKQVAPVETILL  
AEEPKEPKEWYKPSNQYNQELPATGSSKLISAGKQKLTPLAEVLIKYPDAEAI  
QY 1555 GAGCAATGCAATATTCTTGTTCATGCACTCATATGATGGCTTGGTAATGATCCG 1614  
Db 259431 GAGCCGATGAAATCATATATATACACGGTTTATATGATGATGATTAGCATGCAACT 259490  
QY 1615 CTTAGTCATGCTGCTGAATAATTGGGTATGATGATCAACCAATTCATGCTCACTGCTC 1674  
Db 259491 CTTAGCTTGGCTCTTAAATAATTGGGTATGATGATCAACCAACGCTCACTTCAACAGGCTC 259550  
QY 1675 GCTATGATGATGAGGCCATTTTTCGCGCGCTTGAACGATGCTCATCTCATGCTCGCT 1734  
Db 259551 AATATGATGAGAAAAAGTTTTCATATCATAGACAAAGCGCTCAGACGAGCGGG 259610  
QY 1735 AATGCTTTAGTCGACGACATTTTGGCGGATTTGGTATCAAAAGTTATCTAGAAATGCGG 1794  
Db 259611 AATGCTTGGTGGCCATATGATTTGGAGAGGCTTATGATCAAAACATTAATCTTGCCTACGT 259670  
QY 1795 GCAACCGTCTGTGAAACCCCTCTCCATGTCGTCGCGCATCGGCTCACTTTGCAAGAGCT 1854  
Db 259671 CGGCTTACCTTTCAGCAAAATCTCATATGATGATGGCTTACCGGCTCAAAAGGCGCT 259730  
QY 1855 TCCATGTCATATAAATTGACGATTAAGTTTGGGGTGGCGACTAGTAAATTCAGCAGAA 1914  
Db 259731 TCCATGTCATATAAATTGACGATTAAGTTTGGGGTGGCGACTAGTAAATTCAGCAGAA 259790  
QY 1915 TTGGGGTTAAAGAACACGACGACGATCCGCTATCCACAAATCAGGACGATATTGCA 1974  
Db 259791 CATGGGCTCATATGATGATGATGATTTGGGATTTTCTTCMAAGCTTGGCATATTGCA 259850  
QY 1975 GAAACGATACCTTTAAGGCTGCGACGCTTTTACTGCGCGATCACTG--GACTCGAT 2031  
Db 259851 GGCACGATTCATTTGGGTGCAACACCTTCTTACTAGTATGATCAACGACAAATGCTGAT 259910  
QY 2032 GGTACGCTCACTAGAAAGAACCAAAATAGCTGGGATGACATGATATCGGATATCG 2091  
Db 259911 GGGACGCTCAGGTCGAGGAGGACAAAGATATCAGTATGATCAATGTTTGAAGAA 259970  
QY 2092 ACCACTCATACGAGATGCTGTTTATCATCTCCGTCGCGACAAATGACCACTTCT 2151  
Db 259971 AATA-GCCATACGAGATGATATATAGCGGCTATTTCCAGATCAGATGATTCAT 260029  
QY 2152 TCGTATGACCGCTCCGGC 2171  
Db 260030 CAGACATTCATCATTGAGGC 260049  
RESULT 7  
AX023555/c AX023555 4928 bp DNA linear PART 15-SEP-2000  
LOCUS Sequence 10 from Patent WO0003008.  
DEFINITION AX023555  
ACCESSION AX023555  
VERSION AX023555.1 GI:10183916  
KEYWORDS  
SOURCE  
ORGANISM  
artificial sequences.  
REFERENCE  
1 Barth, G., Juretzek, T. and Mauerberger, S.  
AUTHORS Recombinant haploid or diploid yarrowia lipolytica cells for the  
TITLE functional heterologous expression of cytochrome p450 systems  
JOURNAL Patent: WO 0003008-A 10 20-JAN-2000;  
UNIV DRESDEN TECH (DE)  
FEATURES  
source location/Qualifiers  
1..4928  
/organism="synthetic construct"

ORIGIN

Query Match 8.8%; Score 226.4; DB 6; Length 4928;  
Best Local Similarity 83.4%; Pred. No. 1.3e-54;  
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

LTR

2271 ATGAGAGCTTCCAGCAATTAACAATCCGAGCTTACGCTGCGCGCTTAATAA 2330  
2483 ATTAGAGTTACCTGTCGAGATGCTAATCTCCAGAGCGAGTGAACATGAAATTAT 2424  
2331 TTGGCAAGTGTCTGCGCGATACGCTGATGCGCATAGTTAAGCAGCCCGACACCGC 2390  
2423 GGTGCACTCTCAAGTACATCTGCTGATGCGCGCATAGTTAAGCAGCCCGACACCGC 2364  
2391 CAACACCCGCTGACGCGCCCTGATGCGCTTGTCTGCTCCGCGATCCGCTTAAGACAG 2450  
2363 CAACACCCGCTGACGCGCCCTGATGCGCTTGTCTGCTCCGCGATCCGCTTAAGACAG 2304  
2451 CTGTGACCGCTCTCGGAGCGTGCATGTGACAGGTTTCAACGCTCATACCGAAACGG 2510  
2303 CTGTGACCGCTCTCGGAGCGTGCATGTGACAGGTTTCAACGCTCATACCGAAACGG 2244  
2511 CGAGACGAAAGGCGCTGATAGCGCTATTTTATAGTTAATGATGATGATATATG 2570  
2243 CGAGACGAAAGGCGCTGATAGCGCTATTTTATAGTTAATGATGATGATATATG 2184  
2571 TTCTTAG 2578  
2183 TTCTTAG 2176

RESULT 8  
AX023554/c 5391 bp DNA linear PAT 15-SEP-2000  
LOCUS AX023554  
DEFINITION Sequence 9 from Patent WO0003008.  
ACCESSION AX023554  
VERSION AX023554.1 GI:10183915  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Barth, G., Juretzek, T. and Mauersberger, S.  
TITLE Recombinant haploid or diploid yarrowia lipolytica cells for the functional heterologous expression of cytochrome p450 systems  
JOURNAL Patent: WO 0003008-A 9 20-JAN-2000;  
UNIV DRESDEN TECH (DE)  
FEATURES  
Source location/Qualifiers  
1. .5391  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="ringf rmg geschlossene DOUBLEstr ngige DNA, Yarrowia lipolytica/Bescherichia coli-DNA Plasmid p67zb"  
107. .456  
/note="Zeta-Element des Retrotransposons aus Yarrowia lipolytica. Schmidt-Berger et al. 1994, J Bacteriol 174, 2477-82"  
complement(2904. .3634)  
/note="Zeta-Element des Retrotransposons Y11 aus Yarrowia lipolytica. Schmidt-Berger et al. 1994, J Bacteriol 174, 2477-82"

ORIGIN

Query Match 8.8%; Score 226.4; DB 6; Length 5391;

Best Local Similarity 83.4%; Pred. No. 1.3e-54;  
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

LTR

2271 ATGAGAGCTTCCAGCAATTAACAATCCGAGCTTACGCTGCGCGCTTAATAA 2330  
2342 ATTAGAGTTACCTGTCGAGATGCTAATCTCCAGAGCGAGTGTACATGAAATTAT 2883  
2331 TTGGCAAGTGTCTGCGCGATACGCTGATGCGCATAGTTAAGCAGCCCGACACCGC 2390  
2882 GGTGCACTCTCAAGTACATCTGCTGATGCGCGCATAGTTAAGCAGCCCGACACCGC 2823  
2391 CAACACCCGCTGACGCGCCCTGATGCGCTTGTCTGCTCCGCGATCCGCTTAAGACAG 2450  
2822 CAACACCCGCTGACGCGCCCTGATGCGCTTGTCTGCTCCGCGATCCGCTTAAGACAG 2763  
2451 CTGTGACCGCTCTCGGAGCGTGCATGTGACAGGTTTCAACGCTCATACCGAAACGG 2510  
2762 CTGTGACCGCTCTCGGAGCGTGCATGTGACAGGTTTCAACGCTCATACCGAAACGG 2703  
2511 CGAGACGAAAGGCGCTGATAGCGCTATTTTATAGTTAATGATGATGATATATG 2570  
2702 CGAGACGAAAGGCGCTGATAGCGCTATTTTATAGTTAATGATGATGATATATG 2643  
2571 TTCTTAG 2578  
2642 TTCTTAG 2635

RESULT 9  
AX023552/c 7057 bp DNA linear PAT 15-SEP-2000  
LOCUS AX023552  
DEFINITION Sequence 7 from Patent WO0003008.  
ACCESSION AX023552  
VERSION AX023552.1 GI:10183913  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Barth, G., Juretzek, T. and Mauersberger, S.  
TITLE Recombinant haploid or diploid yarrowia lipolytica cells for the functional heterologous expression of cytochrome p450 systems  
JOURNAL Patent: WO 0003008-A 7 20-JAN-2000;  
UNIV DRESDEN TECH (DE)  
FEATURES  
Source location/Qualifiers  
1. .7057  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="ringf rmg geschlossene DOUBLEstr ngige DNA, Yarrowia lipolytica/Bescherichia coli-DNA Plasmid p67ICUpPro"  
complement(16. .2151)  
/note="Promotor des ICU1-Gens aus Yarrowia lipolytica EMBL-Datenbank X=72848"  
4654. .5384  
/note="Zeta-Element des Retrotransposon Y11 aus Yarrowia lipolytica. Schmidt-Berger et al. 1994, J Bacteriol 174, 2477-82"

ORIGIN

Query Match 8.8%; Score 226.4; DB 6; Length 7057;  
Best Local Similarity 83.4%; Pred. No. 1.4e-54;  
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

LTR

2271 ATGAGAGCTTCCAGCAATTAACAATCCGAGCTTACGCTGCGCGCTTAATAA 2330  
4692 ATTAGAGTTACCTGTCGAGATGCTAATCTCCAGAGCGAGTGTACATGAAATTAT 4633  
2331 TTGGCAAGTGTCTGCGCGATACGCTGATGCGCATAGTTAAGCAGCCCGACACCGC 2390  
4632 GGTGCACTCTCAAGTACATCTGCTGATGCGCGCATAGTTAAGCAGCCCGACACCGC 4573

QY 2391 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTGCTCCCGCATCCGCTTAAGACAG 2450  
DB 4572 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTGCTCCCGCATCCGCTTAAGACAG 4513  
QY 2451 CTGTGACCGCTCCCGGAGCTGATGTGACAGAGTTTTCACCGCATCACGAAAGCG 2510  
DB 4512 CTGTGACCGCTCCCGGAGCTGATGTGACAGAGTTTTCACCGCATCACGAAAGCG 4453  
QY 2511 CGAGACGAAAGGCGCTCGTATACGCTATTTTATAGTTAATGATGATATATATG 2570  
DB 4452 CGAGACGAAAGGCGCTCGTATACGCTATTTTATAGTTAATGATGATATATATG 4393  
QY 2571 TTTCTTAG 2578  
DB 4392 TTTCTTAG 4385

RESULT 10  
AX023551/c  
LOCUS AX023551 9402 bp DNA linear PAT 15-SEP-2000  
DEFINITION Sequence 6 from Patent WO0003008.  
ACCESSION AX023551  
VERSION AX023551.1 GI:10183912  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Barth,G., Juretzek,T. and Mauersberger,S.  
TITLE Recombinant haploid or diploid yarrowia lipolytica cells for the functional heterologous expression of cytochrome p450 systems  
JOURNAL Patent: WO 0003008-A 6 20-JAN-2000;  
UNIV DRESDEN TECH (DE)

FEATURES  
Source Location/Qualifiers  
1..9402  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="ringf rmg geschlossene DOUBLER ngise DNA,  
Yarrowia lipolytica/Escherichia coli-DNA Plasmid p671C17a"  
terminator complement(277..9373)  
/note="Terminator des ICL1-Gens aus Yarrowia lipolytica"  
intron complement(1959..2354)  
/note="Intron in ICL1-Gen aus Yarrowia lipolytica  
EMBL-Datenbank X=72848"  
promoter complement(2361..4496)  
/note="Promotor des ICL1-Gens aus Yarrowia lipolytica  
EMBL-Datenbank X=72848"  
LTR 6999..7799  
/note="Zeta-Element aus LTR des Retrotransposon Y11 aus  
Yarrowia lipolytica. Schmidt-Berger et al. 1994, J  
Bacteriol 176,2477-82"

ORIGIN  
Query Match 8.8%; Score 226.4; DB 6; Length 9402;  
Best Local Similarity 83.4%; Pred. No. 1.4e-54;  
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 2271 ATCAGAGCTTCCCATGACATTAACATCCGACCTTAAGCTGCGCTTTAACTAA 2330  
DB 7037 ATTAGAGTACCTGTGCGATGATCACTCTCAAGAGATGATACATGAAATTAT 6978  
QY 2331 TTGGCAAGTGTCTGCGCGATAGCGTATGCGCATAGTTAAGCAAGCCCGACACCG 2390  
DB 6977 GGTGACCTCTCAGTCAATCTGCTGATGCGCATATGTAAGCAAGCCCGACACCG 6918  
QY 2391 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTGCTCCCGCATCCGCTTAAGACAG 2450  
DB 6917 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTGCTCCCGCATCCGCTTAAGACAG 6858  
QY 2451 CTGTGACCGCTCCCGGAGCTGATGTGACAGAGTTTTCACCGCATCACGAAAGCG 2510  
DB 6857 CTGTGACCGCTCCCGGAGCTGATGTGACAGAGTTTTCACCGCATCACGAAAGCG 6798

QY 2511 CGAGACGAAAGGCGCTCGTATACGCTATTTTATAGTTAATGATGATATATATG 2570  
DB 6797 CGAGACGAAAGGCGCTCGTATACGCTATTTTATAGTTAATGATGATATATATG 6738  
QY 2571 TTTCTTAG 2578  
DB 6737 TTTCTTAG 6730

RESULT 11  
AX023553/c  
LOCUS AX023553 10890 bp DNA linear PAT 15-SEP-2000  
DEFINITION Sequence 8 from Patent WO0003008.  
ACCESSION AX023553  
VERSION AX023553.1 GI:10183914  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Barth,G., Juretzek,T. and Mauersberger,S.  
TITLE Recombinant haploid or diploid yarrowia lipolytica cells for the functional heterologous expression of cytochrome p450 systems  
JOURNAL Patent: WO 0003008-A 8 20-JAN-2000;  
UNIV DRESDEN TECH (DE)

FEATURES  
Source Location/Qualifiers  
1..10890  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="ringf rmg geschlossene DOUBLER ngise DNA,  
Yarrowia lipolytica/Escherichia coli-DNA Plasmid p671L43"  
terminator complement(3..274)  
/note="Terminator des ICL1-Gens aus Yarrowia lipolytica  
EMBL-Datenbank X=72848"  
intron complement(3475..3801)  
/note="Intron in ICL1-Gen aus Yarrowia lipolytica  
EMBL-Datenbank X=72848"  
promoter complement(3801..5983)  
/note="Promotor des ICL1-Gens aus Yarrowia lipolytica  
EMBL-Datenbank X=72848"  
LTR 8486..9216  
/note="Zeta-Element des Retrotransposons Y11 aus Yarrowia  
lipolytica. Schmidt-Berger et al. 1994, J Bacteriol 174,  
2477-82"

ORIGIN  
Query Match 8.8%; Score 226.4; DB 6; Length 10890;  
Best Local Similarity 83.4%; Pred. No. 1.5e-54;  
Matches 257; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 2271 ATCAGAGCTTCCCATGACATTAACATCCGACCTTAAGCTGCGCTTTAACTAA 2330  
DB 8524 ATTAGAGTACCTGTGCGATGATCACTCTCAAGAGATGATACATGAAATTAT 8465  
QY 2331 TTGGCAAGTGTCTGCGCGATAGCGTATGCGCATAGTTAAGCAAGCCCGACACCG 2390  
DB 8464 GTGACCTCTCAGTCAATCTGCTGATGCGCATATGTAAGCAAGCCCGACACCG 8405  
QY 2391 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTGCTCCCGCATCCGCTTAAGACAG 2450  
DB 8404 CAACACCCGCTGACGGCCCTGACGGGCTTGTCTGCTCCCGCATCCGCTTAAGACAG 8345  
QY 2451 CTGTGACCGCTCCCGGAGCTGATGTGACAGAGTTTTCACCGCATCACGAAAGCG 2510  
DB 8344 CTGTGACCGCTCCCGGAGCTGATGTGACAGAGTTTTCACCGCATCACGAAAGCG 8285  
QY 2511 CGAGACGAAAGGCGCTCGTATACGCTATTTTATAGTTAATGATGATATATATG 2570  
DB 8284 CGAGACGAAAGGCGCTCGTATACGCTATTTTATAGTTAATGATGATATATATG 8225  
QY 2571 TTTCTTAG 2578



```

QY      2411 TGAAGGAGCTTGTCTCTCCCGGATCCGCTTACAGACAAGCTGTGACCGCTCTCCGGAGC 2470
      |||
Db      193 TGAAGGAGCTTGTCTCTCCCGGATCCGCTTACAGACAAGCTGTGACCGCTCTCCGGAGC 252
      |||
QY      2471 TGCATGTGTCAAGAGGTTTTCACCGTCATCACGAAACGGCGAGACGAAAGGCGCTCGTG 2530
      |||
Db      253 TGCATGTGTCAAGAGGTTTTCACCGTCATCACGAAACGGCGAGACGAAAGGCGCTCGTG 312
      |||
QY      2531 ATAGCCCTATTATTTTATAGTTAATGTCATGATATATATATGTTTCTTAG 2578
      |||
Db      313 ATAGCCCTATTATTTTATAGTTAATGTCATGATATATATATGTTTCTTAG 360
      |||

```

## RESULT 15

```

AX027819      AX027819      2297 bp      DNA      linear      PAT 16-SEP-2000
LOCUS
DEFINITION      Sequence 11 from Patent WO0039307.
ACCESSION      AX027819
VERSION      AX027819.1 GI:10188663
KEYWORDS
SOURCE
ORGANISM      synthetic construct
              synthetic construct
              artificial sequences.

```

```

REFERENCE
AUTHORS      Besetti, G., Cali, S., Orsini, G., Tonon, G., Zuffi, G. and Ghisotti, D.
TITLE      Recombinant bacterial strains for the production of natural
              nucleosides and modified analogues thereof
JOURNAL      Patent: NO 0039307-A 11 06-JUL-2000;
              BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHEMA SPA (IT) ;
              ORSINI GABRIANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
              GHISOTTI DANIELA (IT)

```

```

FEATURES
source      1..2297
              /organism="synthetic construct"
              /mol_type="unassigned DNA"
              /db_xref="taxon:32630"
              /note="Cloning vector derived from pUC18"

```

## ORIGIN

```

Query Match      8.7%; Score 224.8; DB 6; Length 2297;
Best Local Similarity 99.1%; Pred. No.3.4e-54;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2351 TACGCTGATGCCGATAGTTAAGCCAGGCCCGGACACCCCGGACACCCGCTGACGCGCC 2410
      |||
Db      133 TGCCTGTGATGCCGATAGTTAAGCCAGGCCCGGACACCCCGGACACCCGCTGACGCGCC 192
      |||
QY      2411 TGAAGGAGCTTGTCTCTCCCGGATCCGCTTACAGACAAGCTGTGACCGCTCTCCGGAGC 2470
      |||
Db      193 TGAAGGAGCTTGTCTCTCCCGGATCCGCTTACAGACAAGCTGTGACCGCTCTCCGGAGC 252
      |||
QY      2471 TGCATGTGTCAAGAGGTTTTCACCGTCATCACGAAACGGCGAGACGAAAGGCGCTCGTG 2530
      |||
Db      253 TGCATGTGTCAAGAGGTTTTCACCGTCATCACGAAACGGCGAGACGAAAGGCGCTCGTG 312
      |||
QY      2531 ATAGCCCTATTATTTTATAGTTAATGTCATGATATATATATGTTTCTTAG 2578
      |||
Db      313 ATAGCCCTATTATTTTATAGTTAATGTCATGATATATATATGTTTCTTAG 360
      |||

```

```

Search completed: April 30, 2004, 15:28:51
Job time : 10412.6 secs

```





CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 717 BP; 175 A; 165 C; 187 G; 190 T; 0 U; 0 Other;

XX Query Match 15.8%; Score 408.2; DB 7; Length 717;

XX Best Local Similarity 73.1%; Pred. No. 1.6e-113;

XX Matches 524; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 787 ATGATTTGTCATTCGATATGATTTTGTGCGTCTTGCCGCAATCTATCCAGGTTTGA 846  
 DB 1 ATGATTTGTCATTCGATATGATTTTGTGCGTCTTGCCGCAATCTATCCAGGTTTGA 60  
 QY 847 CTGCTGAAACGCTCTATCCAAATCGGCAAGATTGTGTAATGGGCAATTTGCTGCGG 906  
 DB 61 TTGGTCAACGCAATCCAAACCCCTGATGCGGCTTTGGCGCAATTCGCGTGGT 120  
 QY 907 ATAGTGTATGACGAAGATATGACCGCTCATGCTGAGAAACCTGTGATGAGATTTTGA 966  
 DB 121 TTGGTGTATGATGAGATCTCAACGCAAGGCGGAAACCGCTCAACAAATTTTGA 180  
 QY 967 GCAGCGAGACGATATTTGTGCGCAAAAGTCCATATTCCTATTTTTCAGCGAT 1026  
 DB 181 TCCGCGGTCGCGCATCTGCCCGCAAAATTCATTTACCCCACTTCATCACTGAT 240  
 QY 1027 CCGCTGTTGATGCAACCCCAACGCAATCCGATGCTTGAAGTCCATTTTCCAT 1086  
 DB 241 CTTTGGTGAACGTAATCTTAAACGATCCAAAGCGGTTGAGCATCATTTTCTAC 300  
 QY 1087 TAGCTTTATTAACCGTGGATATTCAAACAAATGAAGATTTTGTATCGAACCCGAG 1146  
 DB 301 TAAGCCCTGCTCATATTCACACGTCACGCAATTTGAAGAGCTGATGATTAAGAC 360  
 QY 1147 CGCGTATATGTTTGTATCTTACTTACTTCAAGAAAGAAATGCGCTGCTTTTGA 1206  
 DB 361 CGTGTATATGTTTGTACATGAGACGATTTTTCAGAGTAACAAATTTTGGCGTTTGA 420  
 QY 1207 CATGTCGCGCAATTCAGATGCGTGGCAAAATTAACGGGCTGCGGTTGAATACATCC 1266  
 DB 421 CATGTCGCGCAATTCAGATGCGTGGCAAAATTAACGGGCTGCGGTTGAATACATCC 480  
 QY 1267 GTGTACTATTTTCAATAGAAAAGTTTGTGAGGAGATTTTATGATGCTTACGCG 1326  
 DB 481 GTGTACTATTTTCAATAGAAAAGTTTGTGAGGAGATTTTATGATGCTTACGCG 540  
 QY 1327 AAAATTTGCGTGAAGTAAATGCAATGACCATTAACGCGCTTGAATCAATACCGGAGTG 1386  
 DB 541 AAAATTTGAGTGAACGTCATGCAATGACCGTAAACGTCGTGATCATACCGGAGTGA 600

QY 1387 ATGCTGACATCAATTAATATGCGCATCTTGTAAAGGCAAGACCAACGACACC 1446  
 DB 601 ATTGTGATGACATTAATTAATGCGCTTACGAAAGTAAAGCGGCAACCTGCGACG 660  
 QY 1447 GTTATGCTGTCGACGATGATGAAGTACCTATTTTCAACCGTTTACAGGTTAA 1503  
 DB 661 GTTATGCTGTCGACGATGATGAAGTACCTATTTTCAACCGTTTACAGGTTAA 717

# RESULT 2

ACAS3529/c  
 ID ACAS3529 standard; DNA; 1308 BP.

ACAS3529;  
 AC 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #35186.

KM Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KM drug design; gene.

OS *Vibrio cholerae*.

PM MO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002MO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

PA (ELIT-1) ELITRA PHARM INC.

PI Wang L, Zumdio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;  
 PI Wal D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.

DR P-PSDB; ABU49559.

PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 41399; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 XX the 6213 antisense sequences given in the specification where expression  
 XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
 XX (1) a vector comprising a promoter operably linked to the nucleic acid  
 XX encoding a polypeptide whose expression is inhibited by the antisense  
 XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 XX polypeptide or its fragment whose expression is inhibited by the  
 XX antisense nucleic acid; (4) an antibody capable of specifically binding  
 XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 XX proliferation or the activity of a gene in an operon required for  
 XX proliferation; (7) identifying a compound that influences the activity of  
 XX the gene product or that has an activity against a biological pathway  
 XX required for proliferation, or that inhibits cellular proliferation; (8)  
 XX identifying a gene required for cellular proliferation or the biological  
 XX pathway in which a proliferation-regulated gene or its gene product lies  
 XX or a gene on which the test compound that inhibits proliferation of an  
 XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 XX compound's activity; (11) a culture comprising strains in which the gene  
 XX product is overexpressed or underexpressed; (12) determining the extent  
 XX to which each of the strains is present in a culture or collection of  
 XX strains; or (13) identifying the target of a compound that inhibits the  
 XX proliferation of an organism. The antisense nucleic acids are useful for  
 XX identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIFO at  
CC ftp.wifo.int/pub/published\_pct\_sequences

XX Sequence 1308 BP; 319 A; 301 C; 324 G; 364 T; 0 U; 0 Other;

Query Match 9.7%; Score 249.2; DB 7; Length 1308;  
Best Local Similarity 70.6%; Pred. No. 8,2e-65;

Matches 332; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

DB 23 ACTTGATGTTACTGGCGGCACTTGTGCTGAGGAGGCTGACGATCTTGAGAG 82  
DB 470 ACCCTCAGGTGTTAAATCAGCAGGACCTCAGCCCAACCTGACGGCTGCTACTTCAAT 411  
QY 83 ATTAAGGCGCATGACGAGGTTTCATTAATAATGCTGCTGCGCACGTAACCTTGATGAGC 142  
DB 410 ACGATCGCCATGACTAGGTTTCATTAATAATGCTGCTGCGCACGTAACCTTGATGAGC 351  
QY 143 ATAGCAGCTGCGGCTTGTGCTGCTGAGGATCCGATTTCAATTGCGCATCGGCT 202  
DB 350 ATCCGAGTGTGCTGTTTCCGCTTGTGACCAATCCCATTTTACTTGTCTTGAAGC 291  
QY 203 TGAATAATGAAATACGCTTAATGCTTAATAATAAGTACATTCATGAGTGCATGCTGC 262  
DB 290 ACGAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 231  
QY 263 TCTAATAATAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 322  
DB 230 GTCATATATATGATATGCGCATCGCAAAAAGCATATGCGCAATGATCTCTTGA 171  
QY 323 CCGTACTTCTCTTAATAATAATGCTTAATGCTTCTCATCTGCGCACATCAATTCATGAGC 382  
DB 170 CGGATGCGCATCAAGAGTCCGCTGCGCATCTTCTCATCTGCGCACATCAATTCATGAGC 111  
QY 383 ACCCTCAATCCGATGATCAAGATGAAAGGCGCTCATATTTGATTTTATAGCATCA 442  
DB 110 ACCGACATCTCGGATTAATAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 51  
QY 443 AGATCAATTAATCTCGGATTAATAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492  
DB 50 AGATGAGAGAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1

RESULT 3  
AB283655  
ID AB283655 standard; cDNA; 528 BP.

XX AB283655;  
XX  
XX 14-MAY-2003 (first entry)

XX Toxicologically relevant human nucleotide sequence #814.

XX Toxicologically relevant gene; toxicological response; gene; ss.

XX Homo sapiens.

XX MO2003016500-A2.

XX 27-FEB-2003.

XX 16-AUG-2002; 2002MO-US026514.

XX 16-AUG-2001; 2001US-0313080P.

XX (PHAS-1) PHASE-1 MOLECULAR TOXICOLOGY INC.

XX Nelt RE, Dunn RT, Adkins K, Pickett G, Kler LD, Schweiser K;

PI Allen P;  
XX WPI; 2003-268322/26.  
DR  
XX Determining a toxicological response to an agent, useful for screening of  
PT drugs, comprises comparing the expression profile of one or more human  
PT toxic response genes to a reference gene expression profile indicative of  
PT toxicity.

XX Claim 1; Page 255; 455pp; English.

XX The present invention describes a method (M1) for determining a  
CC toxicological response to an agent, which comprises comparing the  
CC expression profile of one or more human toxic response genes to a  
CC reference gene expression profile indicative of toxicity, and so  
CC determining the presence of a toxic response to the agent. Also  
CC described: (1) an array comprising one or more polynucleotides selected  
CC from the genes corresponding to the partial sequences given in AB282842  
CC ; and (2) determining if a gene putatively identified to be a toxic  
CC response gene plays a role on toxic response pathways by determining the  
CC expression profile of the gene after exposure of cells or a human subject  
CC to a known toxic pharmaceutical or industrial agent, comprising: (a)  
CC exposing cells to an agent or isolating cells from a human subject who  
CC was exposed to an agent; (b) obtaining the test gene expression profile  
CC for a putatively identified toxic response gene after exposure to a known  
CC toxic pharmaceutical or industrial agent; and (c) comparing the test  
CC profile to the expression profile of a gene with a similar function or  
CC comparing the test profile to the expression profile of that gene after  
CC exposure to other known toxic compounds. The methods are useful for  
CC predicting and determining toxicological responses on a cellular, organ  
CC or system level. The arrays comprising the human genes are useful for  
CC toxicological screening of drugs, pharmaceutical compounds and chemicals  
XX

SO Sequence 528 BP; 124 A; 139 C; 118 G; 144 T; 0 U; 3 Other;

Query Match 8.7%; Score 224.8; DB 7; Length 528;

Best Local Similarity 99.1%; Pred. No. 1.4e-57;

Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCTGATGCGGATGATTAAGCCAGCCCGGACACCCGCAACCCGCTGACGGGCC 2410  
DB 227 TGCTGATGCGGATGATTAAGCCAGCCCGGACACCCGCAACCCGCTGACGGGCC 286  
QY 2411 TGAAGGCTTGTGCTGCTGCGGATCGGCTTACAGCAAGCTGTGACCGTCTCGGAGC 2470  
DB 287 TGAAGGCTTGTGCTGCTGCGGATCGGCTTACAGCAAGCTGTGACCGTCTCGGAGC 346  
QY 2471 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2530  
DB 347 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406  
QY 2531 ATAAGGCTTATTTATAGGTTAATGATGATGATGATGATGATGATGATGATGATGAT 2578  
DB 407 ATAAGGCTTATTTATAGGTTAATGATGATGATGATGATGATGATGATGATGATGAT 454

RESULT 4  
ABK42985  
ID ABK42985 standard; DNA; 878 BP.

XX ABK42985;

XX 21-MAY-2002 (first entry)

XX Genomic sequence #884 encoding novel human connective tissue polypeptide.

XX Human; connective tissue related disorder; cancer; gene therapy;

XX cytostatic; gene; ds.

XX Homo sapiens.

XX MO200155343-A1.

XX 02-AUG-2001.  
 PD 17-JAN-2001; 2001WO-US001322.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-01806528P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218230P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226688P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228824P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 21-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 25-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236337P.

PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249247P.  
 PR 17-NOV-2000; 2000US-0249248P.  
 PR 17-NOV-2000; 2000US-0249249P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256718P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 06-DEC-2000; 2000US-0251480P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 XX

DR WPI; 2001-565190/63.  
XX Nucleic acid encoding novel connective tissue associated polypeptides.  
PT used in diagnosing, preventing, treating or ameliorating a disorder such  
PT as cancer or rheumatoid arthritis.  
PS Disclosure; SEQ ID NO 1872; 673bp; English.  
XX  
XX The present invention relates to the isolation of novel human connective  
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide  
CC (cDNA and genomic) sequences encoding them. The sequences of the  
CC invention are useful in the diagnosis, treatment, prevention and/or  
CC prognosis of diseases associated with connective tissue(s), including  
CC cancer. The polynucleotide sequences of the invention are also useful in  
CC gene therapy. ABR42102-ABR43116 represent genomic sequences encoding the  
CC novel human connective tissue related polypeptides. Note: the sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;  
  
Query Match 8.7%; Score 224.8; DB 4; Length 878;  
Best Local Similarity 99.1%; Pred. No. 1.9e-57;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2351 TACGCTGATGCCGCTAGTTAAGCAGCCCGGACACCCCGCAACCCGCTGAGCGGCC 2410  
DB 202 TCGCTGATGCCGCTAGTTAAGCAGCCCGGACACCCCGCAACCCGCTGAGCGGCC 261  
QY 2411 TGACGGGCTGTGCTGCTCCCGGACCTCCATTACAGACAGCTGACCGTCCGGAGC 2470  
DB 262 TGACGGGCTGTGCTGCTCCCGGACCTCCATTACAGACAGCTGACCGTCCGGAGC 321  
QY 2471 TGACGTGTGACAGGTTTTCACCGTCATCACCAGAACCCCGGAGACAGAGGCTGTG 2530  
DB 322 TGACGTGTGACAGGTTTTCACCGTCATCACCAGAACCCCGGAGACAGAGGCTGTG 381  
QY 2531 ATAGCCCTATTTTATAGTTAATGTCATGATTAATGTTTCTTAG 2578  
DB 382 ATAGCCCTATTTTATAGTTAATGTCATGATTAATGTTTCTTAG 429  
  
RESULT 5  
ID AAS41856 standard; DNA; 878 BP.  
XX  
XX AAS41856;  
AC  
DT 17-DEC-2001 (first entry)  
XX  
XX Genomic sequence #172 encoding novel human enzyme polypeptide.  
DE  
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KM ligase; hyperproliferative disorder; immunodeficiency disorder;  
KM autoimmune disorder; neurological disorder; metabolic disorder;  
KM inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KM blood-related disorder; infectious disorder; gene therapy; cytostatic;  
XX anti arthritic; nephrotropic; anticoagulant; de.  
XX  
OS Homo sapiens.  
XX  
XX WO20015301-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001239.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214868P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 11-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227189P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234233P.  
PR 21-SEP-2000; 2000US-0234747P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 25-SEP-2000; 2000US-0234984P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 27-SEP-2000; 2000US-0236277P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.





CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (1) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;

Query Match 8.7%; Score 224.8; DB 4; Length 878;  
Best Local Similarity 99.1%; Pred. No. 1.9e-57;

Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCGCCGATGTTAGCCAGCCGCGACACCCGCGACACCCGCTGACGCGCC 2410  
DB 202 TGCCTGATGCGCGATGTTAGCCAGCCGCGACACCCGCGACACCCGCTGACGCGCC 261  
QY 2411 TGAAGGGCTTGTCTGCTCCGCGATCCGCTTACAGACAGCTGTACCGCTCCGCGAGC 2470  
DB 262 TGAAGGGCTTGTCTGCTCCGCGATCCGCTTACAGACAGCTGTACCGCTCCGCGAGC 321  
QY 2471 TGCATGTGTCAGAGGTTTACCGTCATCACCGAAGCGCGAGACGAAAGGCTCTG 2530  
DB 322 TGCATGTGTCAGAGGTTTACCGTCATCACCGAAGCGCGAGACGAAAGGCTCTG 381  
QY 2531 ATACGGCTATTTTATAGGTTATGATGATATATAGGTTCTAG 2578  
DB 382 ATACGGCTATTTTATAGGTTATGATGATATATAGGTTCTAG 429

RESULT 7  
AA07341  
ID AA07341 standard; DNA; 878 BP.

XX AA07341;  
XX  
DT 21-NOV-2001 (first entry)

XX Human reproductive system related antigen DNA SEQ ID NO: 10029.

KW Human; reproductive system related antigen; reproductive system disorder;  
cancer; gene therapy; ds.

OS Homo sapiens.  
XX  
XX PN WO200155320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001339.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 19-MAY-2000; 2000US-0198123P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216477P.

XX 11-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234224P.  
PR 25-SEP-2000; 2000US-0234987P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236377P.  
PR 29-SEP-2000; 2000US-0236378P.  
PR 29-SEP-2000; 2000US-0236379P.  
PR 29-SEP-2000; 2000US-0236380P.  
PR 29-SEP-2000; 2000US-0236381P.  
PR 29-SEP-2000; 2000US-0236382P.  
PR 29-SEP-2000; 2000US-0236383P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 13-OCT-2000; 2000US-0239938P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241222P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.



PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
XX  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
XX  
XX  
XX Disclosure; SEQ ID NO 10029; 1297pp + Sequence Listing; English.  
XX  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders,  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention  
XX  
XX  
XX Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;  
XX  
XX  
XX Query Match 8.7%; Score 224.8; DB 4; Length 878;  
XX Best Local Similarity 99.1%; Pred. No. 1.9e-57;  
XX Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2471 TGCATGTTCAGAGGTTTTCACCCGTCATCA CCGAAA GCGCGAGACGAAAGGCTCGTG 2530  
Db 322 TGCATGTTCAGAGGTTTTCACCCGTCATCA CCGAAA GCGCGAGACGAAAGGCTCGTG 381  
Qy 2531 ATACGCTATTATTATAGCTATATGTCAGATATAATGGTTCTTAG 2578  
Db 382 ATACGCTATTATTATAGCTATATGTCAGATATAATGGTTCTTAG 429  
RESULT 8  
ID AAL07345 standard; DNA; 878 BP.  
XX AAL07345;  
AC  
XX  
XX 21-NOV-2001 (first entry)  
DT  
XX  
XX Human reproductive system related antigen DNA SEQ ID NO: 10033.  
DE  
XX Human reproductive system related antigen; reproductive system disorder;  
XX cancer; gene therapy; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200155320-A2.  
XX  
XX  
XX 02-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001WO-08001339.  
XX  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 14-JUL-2000; 2000US-0217496P.  
XX 26-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226688P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.

[illegible]

Query	Match	8.7%	Score 224.8	DB 4	Length 878
Beet Local Similarity	99.1%	Pred. No. 1.9e-57			
Matches 226	Conservative 0	Mismatches 2	Indels 0	Gaps 0	
2351	TACGCTGATGCCGATAGTTAAGCCAGCCCGGACACCCCGCCACACCCCGCTGACGCGCC	2410			
202	TGCTCTGATGCCGATAGTTAAGCCAGCCCGGACACCCCGCCACACCCCGCTGACGCGCC	261			
2411	TGACGGGGTTTCTCTCCCGGCACTCCGCTTACAGAGAAGCTGACCGCTCCGGAGC	2470			
262	TGACGGGGTTTCTCTCCCGGCACTCCGCTTACAGAGAAGCTGACCGCTCCGGAGC	321			
2471	TGCACTGTCTCAGAGGTTTTCACCGTCTACACCGAAACCGCGGAGACGAAAGGCGCTCGT	2530			
322	TGCACTGTCTCAGAGGTTTTCACCGTCTACACCGAAACCGCGGAGACGAAAGGCGCTCGT	381			
2531	ATAGCCCTATTTTATAGCTTAAGTCAATGATGATTAATGTTCTTTG	2578			
382	ATAGCCCTATTTTATAGCTTAAGTCAATGATGATTAATGTTCTTTG	429			

KW	immunomodulator; cardiovascular; cytosolic; nephrothropic;	PR	14-SEP-2000; 2000US-0233063P.
KW	cardiovascular; autoimmune disease; rheumatoid arthritis;	PR	14-SEP-2000; 2000US-0233064P.
KW	hyperproliferative disorder; breast neoplasm; cancer;	PR	14-SEP-2000; 2000US-0233065P.
KW	cardiovascular disorder; cardiac arrest; cerebrovascular disorder;	PR	21-SEP-2000; 2000US-0234223P.
KW	cerebral ischemia; angiogenesis; nervous system disorder;	PR	21-SEP-2000; 2000US-0234274P.
KW	Alzheimer's disease; infection; ocular disorder; corneal infection;	PR	25-SEP-2000; 2000US-0234977P.
KW	wound healing; epithelial cell proliferation; food additive.	PR	25-SEP-2000; 2000US-0234988P.
OS	Homo sapiens.	PR	26-SEP-2000; 2000US-0235484P.
PN	WO200155312-A2.	PR	27-SEP-2000; 2000US-0235834P.
XX		PR	27-SEP-2000; 2000US-0235836P.
XX		PR	29-SEP-2000; 2000US-0236327P.
PD	02-AUG-2001.	PR	29-SEP-2000; 2000US-0236367P.
XX		PR	29-SEP-2000; 2000US-0236368P.
XX		PR	29-SEP-2000; 2000US-0236369P.
PF	17-JAN-2001; 2001WO-US001321.	PR	29-SEP-2000; 2000US-0236370P.
XX		PR	02-OCT-2000; 2000US-0236802P.
XX		PR	02-OCT-2000; 2000US-0237037P.
PR	31-JAN-2000; 2000US-0179065P.	PR	02-OCT-2000; 2000US-0237038P.
PR	04-FEB-2000; 2000US-0180628P.	PR	02-OCT-2000; 2000US-0237039P.
PR	24-FEB-2000; 2000US-0184664P.	PR	02-OCT-2000; 2000US-0237040P.
PR	02-MAR-2000; 2000US-0186350P.	PR	13-OCT-2000; 2000US-0239835P.
PR	16-MAR-2000; 2000US-0188874P.	PR	13-OCT-2000; 2000US-0239837P.
PR	17-MAR-2000; 2000US-0190076P.	PR	20-OCT-2000; 2000US-0240960P.
PR	18-APR-2000; 2000US-0198123P.	PR	20-OCT-2000; 2000US-0241211P.
PR	19-MAY-2000; 2000US-0205515P.	PR	20-OCT-2000; 2000US-0241785P.
PR	07-JUN-2000; 2000US-0209467P.	PR	20-OCT-2000; 2000US-0241787P.
PR	28-JUN-2000; 2000US-0214886P.	PR	20-OCT-2000; 2000US-0241808P.
PR	30-JUN-2000; 2000US-0215135P.	PR	20-OCT-2000; 2000US-0241809P.
PR	07-JUL-2000; 2000US-0216647P.	PR	20-OCT-2000; 2000US-0241826P.
PR	07-JUL-2000; 2000US-0216880P.	PR	01-NOV-2000; 2000US-0244617P.
PR	11-JUL-2000; 2000US-0217487P.	PR	08-NOV-2000; 2000US-0246474P.
PR	14-JUL-2000; 2000US-0218290P.	PR	08-NOV-2000; 2000US-0246475P.
PR	26-JUL-2000; 2000US-0220963P.	PR	08-NOV-2000; 2000US-0246476P.
PR	26-JUL-2000; 2000US-0220964P.	PR	08-NOV-2000; 2000US-0246477P.
PR	14-AUG-2000; 2000US-0224518P.	PR	08-NOV-2000; 2000US-0246478P.
PR	14-AUG-2000; 2000US-0224519P.	PR	08-NOV-2000; 2000US-0246523P.
PR	14-AUG-2000; 2000US-0225213P.	PR	08-NOV-2000; 2000US-0246524P.
PR	14-AUG-2000; 2000US-0225214P.	PR	08-NOV-2000; 2000US-0246525P.
PR	14-AUG-2000; 2000US-0225267P.	PR	08-NOV-2000; 2000US-0246527P.
PR	14-AUG-2000; 2000US-0225268P.	PR	08-NOV-2000; 2000US-0246528P.
PR	14-AUG-2000; 2000US-0225270P.	PR	08-NOV-2000; 2000US-0246532P.
PR	14-AUG-2000; 2000US-0225477P.	PR	08-NOV-2000; 2000US-0246609P.
PR	14-AUG-2000; 2000US-0225757P.	PR	08-NOV-2000; 2000US-0246610P.
PR	14-AUG-2000; 2000US-0225758P.	PR	08-NOV-2000; 2000US-0246611P.
PR	18-AUG-2000; 2000US-0226273P.	PR	17-NOV-2000; 2000US-0249207P.
PR	22-AUG-2000; 2000US-0226681P.	PR	17-NOV-2000; 2000US-0249208P.
PR	22-AUG-2000; 2000US-0226688P.	PR	17-NOV-2000; 2000US-0249209P.
PR	22-AUG-2000; 2000US-0227182P.	PR	17-NOV-2000; 2000US-0249210P.
PR	23-AUG-2000; 2000US-0227009P.	PR	17-NOV-2000; 2000US-0249211P.
PR	30-AUG-2000; 2000US-0228924P.	PR	17-NOV-2000; 2000US-0249212P.
PR	01-SEP-2000; 2000US-0229287P.	PR	17-NOV-2000; 2000US-0249213P.
PR	01-SEP-2000; 2000US-0229343P.	PR	17-NOV-2000; 2000US-0249214P.
PR	01-SEP-2000; 2000US-0229344P.	PR	17-NOV-2000; 2000US-0249215P.
PR	01-SEP-2000; 2000US-0229345P.	PR	17-NOV-2000; 2000US-0249216P.
PR	05-SEP-2000; 2000US-0229509P.	PR	17-NOV-2000; 2000US-0249217P.
PR	05-SEP-2000; 2000US-0229513P.	PR	17-NOV-2000; 2000US-0249218P.
PR	06-SEP-2000; 2000US-0230437P.	PR	17-NOV-2000; 2000US-0249244P.
PR	06-SEP-2000; 2000US-0230438P.	PR	17-NOV-2000; 2000US-0249245P.
PR	08-SEP-2000; 2000US-0231242P.	PR	17-NOV-2000; 2000US-0249246P.
PR	08-SEP-2000; 2000US-0231243P.	PR	17-NOV-2000; 2000US-0249297P.
PR	08-SEP-2000; 2000US-0231244P.	PR	17-NOV-2000; 2000US-0249297P.
PR	08-SEP-2000; 2000US-0231413P.	PR	17-NOV-2000; 2000US-0249299P.
PR	08-SEP-2000; 2000US-0231414P.	PR	01-DEC-2000; 2000US-0250160P.
PR	08-SEP-2000; 2000US-0230801P.	PR	01-DEC-2000; 2000US-0250301P.
PR	12-SEP-2000; 2000US-0231968P.	PR	05-DEC-2000; 2000US-0251030P.
PR	14-SEP-2000; 2000US-0232397P.	PR	05-DEC-2000; 2000US-0251988P.
PR	14-SEP-2000; 2000US-0232398P.	PR	05-DEC-2000; 2000US-0251799P.
PR	14-SEP-2000; 2000US-0232399P.	PR	08-DEC-2000; 2000US-0251856P.
PR	14-SEP-2000; 2000US-0232400P.		
PR	14-SEP-2000; 2000US-0232401P.		



Query Match 8.7%; Score 224.8; DB 7; Length 878;  
Best Local Similarity 99.1%; Pred. No. 1.9e-57;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
DB 2351 TACGCTGATGCCGATAGTTAAGCCAGCCGACCCGACACCCGCTGAGCGGCC 2410  
202 TGCTCTGATGCCGATAGTTAAGCCAGCCGACCCGACACCCGCTGAGCGGCC 261  
QY 2411 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGCTCCGGAGC 2470  
DB 262 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGCTCCGGAGC 321  
QY 2471 TGACATGTGTACAGGCTTTTCAACCGTCATACCGAAGCGCGGACAGAAAGGCGCTG 2530  
DB 322 TGACATGTGTACAGGCTTTTCAACCGTCATACCGAAGCGCGGACAGAAAGGCGCTG 381  
QY 2531 ATAGCCCTATTTTATAGTTATATGTCATGATATATATGTTCTTAG 2578  
DB 382 ATAGCCCTATTTTATAGTTATATGTCATGATATATATGTTCTTAG 429

## RESULT 11

ADA41575 standard; DNA; 878 BP.

XX ADA41575;  
XX 20-NOV-2003 (first entry)  
XX Human secreted protein related DNA.  
XX Human; secreted protein; cancer; hyperproliferative disorder;  
XX rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;  
XX anaemia; allergic reaction; asthma; cardiovascular disorder;  
XX wound healing; cytostatic; immunosuppressive; nocrotropic; neuroprotective;  
XX antiviral; anti-allergic; hepatocrotropic; antidiabetic; antiinflammatory;  
XX vulnerrary; cardiac; gene therapy; ds.  
XX Homo sapiens.  
XX WO2002102993-A2.  
XX 27-DEC-2002.  
XX 19-MAR-2002; 2002WO-US008123.  
XX 21-MAR-2001; 2001US-0277340P.  
XX 19-JUL-2001; 2001US-0306171P.  
XX 13-NOV-2001; 2001US-0331287P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX WPI; 2003-175238/17.  
XX New human secreted proteins and nucleic acid molecules, useful for  
XX preparing a diagnostic or pharmaceutical composition for diagnosing,  
XX preventing or treating cancer or other hyperproliferative disorder,  
XX asthma, allergies or AIDS.  
XX Disclosure; SEQ ID NO 1958; 3205bp; English.  
XX  
XX The invention relates to novel genes ADA39629-ADA0565 and proteins  
XX ADA0566-ADA41501 for human secreted proteins, useful for preventing,  
XX treating or ameliorating medical conditions e.g. by protein or gene  
XX therapy. The polypeptides, nucleic acid molecules, antibodies or their  
XX fragments, and agonists or antagonists that bind to the polypeptide are  
XX useful for preparing a diagnostic or pharmaceutical composition for  
XX diagnosing or treating cancer or other hyperproliferative disorder. The  
XX polypeptides and nucleic acid molecules are also useful for detecting,  
XX preventing, diagnosing, prognosticating, treating or ameliorating cancer  
XX or other hyperproliferative disorders including neoplasms, autoimmune

CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus  
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic  
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,  
CC thrombocytopenia), allergic reactions including asthma or eczema,  
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory  
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders  
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,  
CC fungal or viral infections including HIV/AIDS), or wound healing and  
CC disorders of epithelial cell proliferation. The nucleic acids are also  
CC useful for chromosome identification, radiation hybrid mapping or long-  
CC range restriction mapping, as molecular weight markers, or as  
CC hybridization or diagnostic probes. The polypeptides and antibodies are  
CC useful for providing immunological probes for differential identification  
CC of the tissues immunobiochemistry assays. The sequences given in  
CC ADA40803-ADA4165 represent DNA sequences related to human secreted  
CC proteins. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://pub.wipo.int/pub/published_pct_sequences).  
XX

SQ Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;

Query Match 8.7%; Score 224.8; DB 7; Length 878;  
Best Local Similarity 99.1%; Pred. No. 1.9e-57;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCCGATAGTTAAGCCAGCCGACCCGACACCCGCTGAGCGGCC 2410  
DB 202 TGCTCTGATGCCGATAGTTAAGCCAGCCGACCCGACACCCGCTGAGCGGCC 261  
QY 2411 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGCTCCGGAGC 2470  
DB 262 TGACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGCTCCGGAGC 321  
QY 2471 TGACATGTGTACAGGCTTTTCAACCGTCATACCGAAGCGCGGACAGAAAGGCGCTG 2530  
DB 322 TGACATGTGTACAGGCTTTTCAACCGTCATACCGAAGCGCGGACAGAAAGGCGCTG 381  
QY 2531 ATAGCCCTATTTTATAGTTATATGTCATGATATATATGTTCTTAG 2578  
DB 382 ATAGCCCTATTTTATAGTTATATGTCATGATATATATGTTCTTAG 429

## RESULT 12

ACC50906 standard; cDNA; 878 BP.

XX ACC50906;  
XX 12-JUN-2003 (first entry)  
XX Human secreted protein BAC clone SEQ ID NO 1086.  
XX Cardiac; antiarrhythmic; antiarteriosclerotic; vasotrophic; cytostatic;  
XX vulnerable; antiinflammatory; nocrotropic; neuroprotective;  
XX antiparkinsonian; gene therapy; human; cardiovascular disorder; ss.  
XX Homo sapiens.  
XX WO200295010-A2.  
XX 28-NOV-2002.  
XX 19-MAR-2002; 2002WO-US009785.  
XX 21-MAR-2001; 2001US-0277340P.  
XX 19-JUL-2001; 2001US-0306171P.  
XX 13-NOV-2001; 2001US-0331287P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX

DR WPI; 2003-129429/12.  
XX Novel human secreted proteins, useful for detecting, preventing,  
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular  
PT disorders such as arrhythmia.  
XX  
XX Disclosure; SEQ ID NO 1086; 1881bp; English.  
XX  
XX The present invention relates to novel human secreted proteins (ABR47633-  
CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins  
CC and their coding sequences are useful for the preparation of a diagnostic  
CC or pharmaceutical composition for diagnosing or treating a cardiovascular  
CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary  
CC arteriosclerosis and myocardial ischemia), neural disorders, immune  
CC system disorders, muscular disorders, reproductive disorders,  
CC gastrointestinal disorders, pulmonary disorders, renal disorders,  
CC proliferative disorders and/or cancerous diseases and conditions, for  
CC wound healing and epithelial cell proliferation, to treat inflammation or  
CC infection, for treating thrombosis and arteriosclerosis, for treating or  
CC preventing neural damage which occurs in neuronal disorders or  
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's  
CC disease, to enhance bone and periodontal regeneration and aid in tissue  
CC transplants or bone grafts, to prevent skin aging or hair loss, to  
CC stimulate growth and differentiation of hematopoietic cells and bone  
CC marrow cells when used in combination with other cytokines, to maintain  
CC organs before transplantation or for supporting cell culture of primary  
CC tissues, to increase or decrease differentiation or proliferation of  
CC embryonic stem cells, or to modulate mammalian characteristics or  
CC metabolism. The present sequence was used to illustrate the invention.  
CC Note: The sequence data for this patent was published in electronic  
CC format and is available from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;  
SQ  
Query Match 8.7%; Score 224.8; DB 7; Length 878;  
Best Local Similarity 99.1%; Pred. No. 1.9e-57;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TAGCGTATGTCGCGATGTTAAGTAAAGCCAGCCCGCAACCCCGCTGACGGCC 2410  
DB 202 TGCTCTGATGCCGATAGTTAAAGCCAGCCCGCAACCCCGCTGACGGCC 261  
QY 2411 TGACGGGCTGTGCTGCTCCCGGATCGCTTACAGCAAGCTGTGACCGTCCGGAGC 2470  
DB 262 TGACGGGCTGTGCTGCTCCCGGATCGCTTACAGCAAGCTGTGACCGTCCGGAGC 321  
QY 2471 TGACGTGTGAGAGGTTTTCACCGTCATCAACCGAAAGCGGAGAAAGGCGCTG 2530  
DB 322 TGACGTGTGAGAGGTTTTCACCGTCATCAACCGAAAGCGGAGAAAGGCGCTG 381  
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 2578  
DB 382 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 429

## RESULT 13

ABZ71509

ID ABZ71509 standard; DNA; 878 BP.

XX ABZ71509;

DT 03-APR-2003 (first entry)

DE Secreted protein gene 74 genomic fragment HMOG30, SEQ ID NO:619.  
XX  
XX Human; secreted protein; digestive disorder; gastrointestinal disorder;  
XX mouth; oesophagus; stomach; small intestine; large intestine; liver;  
XX biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;  
XX immune disorder; inflammation; infection; wound healing; drug screening;  
XX chromosome identification; chromosome mapping; cytostatic; gene therapy;  
XX antiinflammatory; immunosuppressive; valvular; chromosome 15; gene; ds.

OS Homo sapiens.  
XX  
XX WO200276488-A1.  
XX  
XX 03-OCT-2002.  
XX  
XX 19-MAR-2002; 2002WO-US008276.  
XX  
XX 21-MAR-2001; 2001US-0277340P.  
XX 19-JUL-2001; 2001US-0306171P.  
XX 13-NOV-2001; 2001US-0331287P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM,  
XX  
XX WPI; 2003-029900/02.  
XX  
XX New human secreted proteins and nucleic acids, useful for detecting,  
PT preventing, diagnosing, prognosticating, treating and/or ameliorating  
PT e.g. gastrointestinal diseases and disorders, or cancers.  
XX  
XX Disclosure; Page 1184; 1216bp; English.  
XX  
XX ABR71190-ABZ71478 represent cDNAs corresponding to 178 human secreted  
CC protein genes, and ABR00011-ABP00299 represent the proteins they encode.  
CC ABR71479-ABZ71540 represent human secreted protein genomic fragments. The  
CC invention also encompasses antibodies specific for the secreted proteins,  
CC the use of the secreted proteins in drug screening, and recombinant  
CC vectors and host cells comprising a nucleic acid of the invention. The  
CC secreted proteins, nucleic acids encoding them, antibodies or antibody  
CC fragments specific for the secreted proteins, and modulators of protein  
CC activity are useful for diagnosing, treating, ameliorating or preventing  
CC digestive disorders. Such conditions include disorders of the mouth,  
CC oesophagus, stomach, small intestine, large intestine, liver, biliary  
CC tract and pancreas, and include cancers of these organs and tissues. The  
CC secreted proteins and their nucleic acids may also be used in the  
CC treatment of immune disorders, inflammation, infection,  
CC hyperproliferative disorders, and to promote wound healing. Nucleic acids  
CC of the invention may be used for chromosome identification, chromosome  
CC mapping, in gene therapy, for identifying individuals from minute  
CC biological samples, as hybridisation probes, and as molecular weight  
CC markers. The present sequence represents a human secreted protein genomic  
CC fragment referred to in the disclosure of the invention  
XX  
XX Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;  
SQ  
Query Match 8.7%; Score 224.8; DB 7; Length 878;  
Best Local Similarity 99.1%; Pred. No. 1.9e-57;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TAGCGTATGTCGCGATGTTAAGTAAAGCCAGCCCGCAACCCCGCTGACGGCC 2410  
DB 202 TGCTCTGATGCCGATAGTTAAGCCAGCCCGCAACCCCGCTGACGGCC 261  
QY 2411 TGACGGGCTGTGCTGCTCCCGGATCGCTTACAGCAAGCTGTGACCGTCCGGAGC 2470  
DB 262 TGACGGGCTGTGCTGCTCCCGGATCGCTTACAGCAAGCTGTGACCGTCCGGAGC 321  
QY 2471 TGACGTGTGAGAGGTTTTCACCGTCATCAACCGAAAGCGGAGAAAGGCGCTG 2530  
DB 322 TGACGTGTGAGAGGTTTTCACCGTCATCAACCGAAAGCGGAGAAAGGCGCTG 381  
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 2578  
DB 382 ATAGCGCTATTTTATAGTTAATGTCATGATATATATGTTCTTAG 429

## RESULT 14

ADB91870

ID ADB91870 standard; DNA; 878 BP.

XX ADB91870;

XX 04-DEC-2003 (first entry)  
 DT Human secreted protein related DNA #SEQ ID 816.  
 XX  
 DE Secreted protein; gene therapy; antidiabetic; diabetes; human; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003004622-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 19-MAR-2002; 2002WO-US008124.  
 XX  
 PR 21-MAR-2001; 2001US-0277340P.  
 PR 19-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 DR WPI; 2003-229407/22.  
 XX  
 PT Nucleic acid encoding a human secreted protein is useful in diagnosing or  
 XX treating diabetes or conditions related to diabetes.  
 PS Disclosure; SEQ ID NO 816; 1537bp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules ADB91065-  
 CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-  
 CC ADB91834. Also disclosed is a recombinant vector comprising a  
 CC polynucleotide of the invention, and a recombinant host cell comprising  
 CC the recombinant vector. The polypeptide of the invention is useful in  
 CC identifying a binding partner by contacting the polypeptide with a  
 CC binding partner, and determining whether the binding partner increases or  
 CC decreases activity of the polypeptide. The polypeptide, polynucleotide,  
 CC antibody or its fragment, agonist or antagonist are useful for preparing  
 CC a pharmaceutical composition for diagnosing or treating diabetes or  
 CC conditions related to diabetes. The present sequence is that of the human  
 CC immunoglobulin Fc portion used to generate fusion proteins, increasing  
 CC the stability of the fused protein as compared to the secreted protein  
 CC only. Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 878 BP; 213 A; 229 C; 199 G; 236 T; 0 U; 1 Other;  
 XX  
 Query Match 8.7%; Score 224.8; DB 8; Length 878;  
 Best Local Similarity 99.1%; Pred. No. 1.9e-57;  
 Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 QY 2351 TACGCTGATGCGCATGTTAAGCAGAGCCCGGACAGCCCGGATGAGCGGCC 2410  
 DB 202 TGCTTGATGCGCATGTTAAGCAGAGCCCGGACAGCCCGGATGAGCGGCC 261  
 QY 2411 TGACGCGCTTGTCTCTCCCGCATCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470  
 DB 262 TGACGCGCTTGTCTCTCCCGCATCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 321  
 QY 2471 TGACGCTGTCAGAGTTTTCACCGTCATCCGAAAGCGGAGAGCGAGCGCTGTTG 2510  
 DB 322 TGACGCTGTCAGAGTTTTCACCGTCATCCGAAAGCGGAGAGCGAGCGCTGTTG 381  
 QY 2531 ATACGCTATTTTATAGATTATGTCATGATTAATATGTTCTTAG 2578  
 DB 382 ATACGCTATTTTATAGATTATGTCATGATTAATATGTTCTTAG 429  
 XX  
 RESULT 15  
 ADB61141  
 ID ADB61141 standard; DNA; 878 BP.

XX ADB61141;  
 AC  
 XX 04-DEC-2003 (first entry)  
 DT  
 XX  
 DE Connective tissue related genomic DNA #884.  
 XX  
 CC Cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiovascular;  
 KW antiarteriosclerotic; immunosuppressive; antirheumatic; antiarrhythmic;  
 KW antiinflammatory; antiallergic; antiaesthetic; dermatological;  
 KW nephrotoxic; virucide; fungicide; antibacterial; antiparasitic;  
 KW gene therapy; ds; connective tissues disorder; rheumatoid arthritis;  
 KW systemic lupus erythematosus; scleroderma; Sjogren's syndrome; cancer;  
 KW cancer metastasis; neoplasia; leukemia; neurodegenerative disorder;  
 KW Alzheimer's disease; Parkinson's disease; cardiovascular disease;  
 KW atherosclerosis; myocarditis; cardiopulmonary bypass complication;  
 KW autoimmune disease; multiple sclerosis; allergic reaction; asthma;  
 KW rhinitis; eczema; inflammatory condition; Crohn's disease; nephritis;  
 KW gastrointestinal disorder; inflammatory bowel disease;  
 KW organ transplant rejection; immune system disorder; Bruton's disease;  
 KW X-linked lymphoproliferative syndrome;  
 KW B-cell lymphoproliferative disorder; HIV; AIDS; infection;  
 KW chromosome identification; chromosome mapping;  
 KW connective tissue related polynucleotide; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003054375-A1.  
 XX  
 PD 20-MAR-2003.  
 XX  
 PF 07-MAR-2002; 2002US-00092154.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225477P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 14-AUG-2000; 2000US-0226279P.  
 PR 18-AUG-2000; 2000US-0226281P.  
 PR 22-AUG-2000; 2000US-0226688P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.





Mon May 3 09:19:46 2004

us-10-603-260-1.rng

Page 17

Search completed: April 30, 2004, 11:04:10  
Job time : 1064.2 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:27 ; Search time 7318.21 Seconds  
(without alignments)  
10519.601 Million cell updates/sec

Title: US-10-603-260-1  
Perfect score: 2578  
Sequence: 1 agctgcacttaacgacca.....tgataataagttctcttag 2578

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estbam:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estropl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estin:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vtl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	227	8.8	495	29	CC818374 100004B07
2	225.4	8.7	496	29	CC818523
3	225.4	8.7	503	29	CC818524 100004L13
4	224.8	8.7	308	14	CC818524 100006N08
					CF099977 rv37e09.y

5	224.8	8.7	364	29	CC818574
6	224.8	8.7	374	29	CC820036
7	224.8	8.7	414	29	CC819240
8	224.8	8.7	464	29	CC816935
9	224.8	8.7	471	29	CC817727
10	224.8	8.7	491	29	CC819923
11	224.8	8.7	507	9	AJ568083
12	224.8	8.7	515	29	CC817752
13	224.8	8.7	519	29	CC817112
14	224.8	8.7	519	29	CC817112
15	224.8	8.7	519	29	CC817796
16	224.8	8.7	521	29	CC819067
17	224.8	8.7	527	29	CC819808
18	224.8	8.7	533	29	CC819841
19	224.8	8.7	539	29	CC818325
20	224.8	8.7	541	29	CC817082
21	224.8	8.7	541	29	CC819843
22	224.8	8.7	542	29	CC816892
23	224.8	8.7	542	29	CC819005
24	224.8	8.7	548	29	CC817171
25	224.8	8.7	551	29	CC816905
26	224.8	8.7	554	29	CC819058
27	224.8	8.7	559	29	CC817875
28	224.8	8.7	563	29	CC819270
29	224.8	8.7	565	29	CC817022
30	224.8	8.7	566	29	CC820024
31	224.8	8.7	567	29	CC817070
32	224.8	8.7	568	29	CC818640
33	224.8	8.7	571	29	CC818423
34	224.8	8.7	580	29	CC819098
35	224.8	8.7	581	29	TA228020
36	224.8	8.7	583	29	CC817633
37	224.8	8.7	584	29	CC818436
38	224.8	8.7	584	29	CC819862
39	224.8	8.7	584	29	CC932326
40	224.8	8.7	586	29	CC816883
41	224.8	8.7	588	29	CC817788
42	224.8	8.7	588	29	CC818340
43	224.8	8.7	589	29	CC817595
44	224.8	8.7	589	29	CC817674
45	224.8	8.7	590	29	CC819754

#### ALIGNMENTS

RESULT 1  
LOCUS CC818374  
DEFINITION 100004B07R Oxytricha plasmid U06C10 library Sterkiella  
histicomscorum genomic clone U06C10004B07 R, genomic survey  
Sequence.  
CC818374 GI:32897661  
GSS.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Sterkiella histicomscorum (Oxytricha trifallax)  
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;  
Stichotrichidae; Oxytrichidae; Sterkiella.

#### REFERENCE

AUTHORS Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
TITLE Paired end reads from plasmid inserts of Oxytricha trifallax  
macronuclear chromosomes  
JOURNAL Unpublished (2003)

#### COMMENT

Contact: Robert B. Weiss  
University of Utah  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0004 row: B column: 07

Seq primer: CACACAGAAACGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 495.  
 Location/Qualifiers  
 1. 495  
 /organism="Sterkiella histriomuscorum"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:94289"  
 /clone="UUGC10004B07"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (G14732114|9B|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.8%; Score 227; DB 29; Length 495;  
 Best Local Similarity 94.0%; Pred. No. 1.2e-58;  
 Matches 236; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

QY 2328 AAATGGCAAGTGTCTGCGGATACGCTGATGCGGATAGTTAAGCCAGCCCGACACC 2387
DB 42 ATATGGGACATCTCAGACATCTGCTGATGCCCATATAGCCAGCCCGACACC 101
QY 2388 CGGCAACACCCGCTGACGCGCCCTGACGGGCTGTGCTCCCGGATCGGCTTACAGAC 2447
DB 102 CGCAACACCCGCTGACGCGCCCTGACGGGCTGTGCTCCCGGATCGGCTTACAGAC 161
QY 2448 AAGCTGTGACCGTCTCCGGAGCTGATGTGTGAGAGGTTTTCACCGTATCAGCGAAG 2507
DB 162 AAGCTGTGACCGTCTCCGGAGCTGATGTGTGAGAGGTTTTCACCGTATCAGCGAAG 221
QY 2508 GCGGAGACGAAAGGCGCTGTGATAGCCTATTTTATAGTTATGTCATGATATTA 2567
DB 222 GCGGAGACGAAAGGCGCTGTGATAGCCTATTTTATAGTTATGTCATGATATTA 281
QY 2568 TGGTTCTTAG 2578
DB 282 TGGTTCTTAG 292

```

RESULT 2 496 bp DNA linear GSS 17-JUL-2003  
 CC818523 100004113 R, genomic survey  
 LOCUS histriomuscorum genomic clone UUGC100004113 R, genomic survey  
 DEFINITION  
 ACCESION CC818523  
 VERSION CC818523.1 GI:32897943  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLES  
 JOURNAL  
 COMMENT  
 1 (bases 1 to 496)  
 Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
 Paired end reads from plasmid inserts of Oxytricha trifallax  
 macronuclear chromosomes  
 Unpublished (2003)  
 Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7277  
 Email: ddu@genetics.utah.edu  
 Plate: 0004 row: L column: 13  
 Seq primer: CACACAGAAACGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 496.  
 Location/Qualifiers  
 1. 496  
 /organism="Sterkiella histriomuscorum"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:94289"  
 /clone="UUGC10004113"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (G14732114|9B|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.7%; Score 225.4; DB 29; Length 496;  
 Best Local Similarity 93.6%; Pred. No. 3.7e-58;  
 Matches 235; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

QY 2328 AAATGGCAAGTGTCTGCGGATACGCTGATGCGGATAGTTAAGCCAGCCCGACACC 2387
DB 41 ATATGGGACATCTCAGACATCTGCTGATGCCCATATAGCCAGCCCGACACC 100
QY 2388 CGGCAACACCCGCTGACGCGCCCTGACGGGCTGTGCTCCCGGATCGGCTTACAGAC 2447
DB 101 CGCAACACCCGCTGACGCGCCCTGACGGGCTGTGCTCCCGGATCGGCTTACAGAC 160
QY 2448 AAGCTGTGACCGTCTCCGGAGCTGATGTGTGAGAGGTTTTCACCGTATCAGCGAAG 2507
DB 161 AAGCTGTGACCGTCTCCGGAGCTGATGTGTGAGAGGTTTTCACCGTATCAGCGAAG 220
QY 2508 GCGGAGACGAAAGGCGCTGTGATAGCCTATTTTATAGTTATGTCATGATATTA 2567
DB 221 GCGGAGACGAAAGGCGCTGTGATAGCCTATTTTATAGTTATGTCATGATATTA 280
QY 2568 TGGTTCTTAG 2578
DB 281 TGGTTCTTAG 291

```

RESULT 3 503 bp DNA linear GSS 17-JUL-2003  
 CC819854 100006088 R, genomic survey  
 LOCUS histriomuscorum genomic clone UUGC100006088 R, genomic survey  
 DEFINITION  
 ACCESION CC819854  
 VERSION CC819854.1 GI:32900533  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLES  
 JOURNAL  
 COMMENT  
 1 (bases 1 to 503)  
 Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
 Paired end reads from plasmid inserts of Oxytricha trifallax  
 macronuclear chromosomes  
 Unpublished (2003)  
 Contact: Robert B. Weiss

University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunne@genetics.utah.edu

Plate: 0006 row: N column: 08  
 Seg primer: CACACGAAACAGCATATAC  
 Class: plasmid ends  
 High quality sequence stop: 503.  
 Location/Qualifiers

## FEATURES

source

1..503  
 /organism="Sterkella histriomuscorum"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:94289"  
 /clone="UUC100006N08"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Oxytricha plasmid UUC10 library"  
 /note="Vector: PWD42nv. Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of PWD42 (g14732114[3b]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.7%; Score 225.4; DB 29; Length 503;  
 Best Local Similarity 93.6%; Pred. No. 3.7e-58;  
 Matches 225; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2328 AATTGGCAAGTGTCTGCGCGATACGTGATCCGATGTAAGCCAGCCCGACACC 2387  
 DB 60 ATATGTGACATCTCAGTACATCTGATGATCCGATGTAAGCCAGCCCGACACC 119  
 QY 2388 CGGCAACCCGCTGACGCGCCCTGACGCGCTTGTCTCCCGATCCGCTTACAGC 2447  
 DB 120 CGCAACACCCGCTGACGCGCCCTGACGCGCTTGTCTCCCGATCCGCTTACAGC 179  
 QY 2448 AAGCTGTGACCGTCTCCGAGCTGATGTCAGAGTTTTCACCGTATACCGAAGC 2507  
 DB 180 AAGCTGTGACCGTCTCCGAGCTGATGTCAGAGTTTTCACCGTATACCGAAGC 239  
 QY 2508 GCGCGAGAGAAAGGCGCTGATACGCTATTTTATAGTTATATGATGATTA 2567  
 DB 240 GCGCGAGAGAAAGGCGCTGATACGCTATTTTATAGTTATATGATGATTA 299  
 QY 2568 TGGTTCTTAG 2578  
 DB 300 TGGTTCTTAG 310

## RESULT 4

CF099977 308 bp mRNA linear EST 22-JUL-2003  
 LOCUS rv37609.y1 Heterodera schachtli J2 PAMP1 v1 Heterodera schachtli  
 DEFINITION cDNA 5', similar to TR:Q51617 Q51617 7.6 KD ORF., mRNA sequence.  
 ACCESSION CF099977  
 VERSION CF099977.1 GI:313139044  
 KEYWORDS EST.  
 SOURCE Heterodera schachtli  
 ORGANISM Heterodera schachtli  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
 Tylenchoidea; Heterodera; Heteroderidae; Heteroderinae; Heterodera.

## REFERENCE

1 (bases 1 to 308)  
 McCarter,J., Clifton,S., Chapelli,B., Page,D., Martin,J.,  
 Wylie,T., Dante,M., Marra,M., Hillier,L., Kucada,T., Theising,B.,

Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,  
 Tsagarishevili,R., Ronko,T., Kennedy,S., Maguire,L., Beck,C.,  
 Underwood,K., Stepien,M., Allen,M., Person,B., Waller,T.,  
 Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,  
 McCann,R., Waterson,R. and Wilson,R.  
 The Washington Univ. Nematode EST Project, 1999  
 Unpublished (1999)  
 Contact: McCarter JP  
 The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

## TITLE

## JOURNAL

## COMMENT

Email: est@watson.wustl.edu  
 The library was constructed by Claire Murphy and Dr. Makedonka Datorova at Washington University, St. Louis  
 (mduco@watson.wustl.edu). The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis kit from Clontech. Directionally cloned into the UDG sites of PAMP1. Heterodera schachtli J2 were provided by Bartel Vanholme of Ghent University (Bartel.Vanholme@ug.ac.be).  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gdbco.  
 Location/Qualifiers

## FEATURES

source

1..308  
 /organism="Heterodera schachtli"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:97005"  
 /dev\_stage="J2"  
 /clone\_lib="Heterodera schachtli J2 PAMP1 v1"  
 /note="Vector: PAMP1, Site 1: Not I; Site 2: Sal I; The library was constructed by Claire Murphy and Dr. Makedonka Datorova at Washington University, St. Louis (mduco@watson.wustl.edu). The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis kit from Clontech. Directionally cloned into the UDG sites of PAMP1. Heterodera schachtli J2 were provided by Bartel Vanholme of Ghent University (Bartel.Vanholme@ug.ac.be)."

## ORIGIN

Query Match 8.7%; Score 224.8; DB 14; Length 308;  
 Best Local Similarity 99.1%; Pred. No. 4.2e-56;  
 Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2251 TAGCTGATGCGGATGTTAAGCAGCCCGACACCGCGGCAACCGGTGAGCGGCC 2410  
 DB 78 TGCTGTATGCGGATGTTAAGCAGCCCGACACCGCGGCAACCGGTGAGCGGCC 137  
 QY 2411 TGACGGGCTTGTCTGCTCCGGGATCCGCTTACAGACAACTGTGACCGGAGC 2470  
 DB 138 TGAAGGCTTGTCTGCTCCGGGATCCGCTTACAGACAACTGTGACCGGAGC 197  
 QY 2471 TGCATGTGTGAGAGTTTTCACGCTATCAACGAAAGCGCGAGAGAGGCGCTGTG 2530  
 DB 198 TGCATGTGTGAGAGTTTTCACGCTATCAACGAAAGCGCGAGAGAGGCGCTGTG 257  
 QY 2531 ATACGCTATTTTATAGTTAATGTCATATATATGTTCTTAG 2578  
 DB 258 ATACGCTATTTTATAGTTAATGTCATATATATGTTCTTAG 305

## RESULT 5

CC818574 364 bp DNA linear GSS 17-JUL-2003  
 LOCUS CC818574  
 DEFINITION 100004E20R Oxytricha plasmid UUC10 library Sterkella  
 histriomuscorum genomic clone UUC100004E20 R, genomic survey  
 sequence.  
 ACCESSION CC818574  
 VERSION CC818574.1 GI:32898040  
 KEYWORDS GSS.

SOURCE  
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)

REFERENCE  
AUTHORS Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
TITLE Paired end reads from plasmid inserts of Oxytricha trifallax  
JOURNAL macronuclear chromosomes  
COMMENT Unpublished (2003)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0004 row: E column: 20  
Seq primer: CACACAGAAACAGCTATGAC  
Class: plasmid ends  
High quality sequence stop: 364.  
Location/Qualifiers

## FEATURES

1. 364  
/organism="Sterkiella histriomuscorum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:94289"  
/clone="U000000420"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repated with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 364;  
Best Local Similarity 99.1%; Pred. No. 4.7e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTATGCGGATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACGGGCC 2410  
DB 87 TGCTGTATGCGGATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACGGGCC 146  
QY 2411 TGAGGGGCTTGTGCTCCCGGATCCGCTTACAGACAGCTGACCGCTCCGGAGC 2470  
DB 147 TGAGGGGCTTGTGCTCCCGGATCCGCTTACAGACAGCTGACCGCTCCGGAGC 206  
QY 2471 TGCATGTGTGAGAGTTTTCACCGTCATCAACGAAAGCGCGAGAGAAAGGCTCTG 2530  
DB 207 TGCATGTGTGAGAGTTTTCACCGTCATCAACGAAAGCGCGAGAGAAAGGCTCTG 266  
QY 2531 ATAGCCCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 2578  
DB 267 ATAGCCCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 314

RESULT 6  
CC820036 374 bp DNA linear GSS 17-JUL-2003  
LOCUS CC820036  
DEFINITION 100005D19R Oxytricha plasmid U0000 library Sterkiella  
histriomuscorum genomic clone U0000006M24 R, genomic survey  
Sequence.  
ACCESSION CC820036  
VERSION CC820036.1 GI:32900898  
KEYWORDS GSS.

SOURCE  
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)

REFERENCE  
AUTHORS Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
TITLE Paired end reads from plasmid inserts of Oxytricha trifallax  
JOURNAL macronuclear chromosomes  
COMMENT Unpublished (2003)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0006 row: M column: 24  
Seq primer: CACACAGAAACAGCTATGAC  
Class: plasmid ends  
High quality sequence stop: 374.  
Location/Qualifiers

## FEATURES

1. 374  
/organism="Sterkiella histriomuscorum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:94289"  
/clone="U0000006M24"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repated with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (g114732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 374;  
Best Local Similarity 99.1%; Pred. No. 4.8e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTATGCGGATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACGGGCC 2410  
DB 92 TGCTGTATGCGGATAGTTAAGCCAGCCCGACACCCGCAACCCGCTGACGGGCC 151  
QY 2411 TGAGGGGCTTGTGCTCCCGGATCCGCTTACAGACAGCTGACCGCTCCGGAGC 2470  
DB 152 TGAGGGGCTTGTGCTCCCGGATCCGCTTACAGACAGCTGACCGCTCCGGAGC 211  
QY 2471 TGCATGTGTGAGAGTTTTCACCGTCATCAACGAAAGCGCGAGAGAAAGGCTCTG 2530  
DB 212 TGCATGTGTGAGAGTTTTCACCGTCATCAACGAAAGCGCGAGAGAAAGGCTCTG 271  
QY 2531 ATAGCCCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 2578  
DB 272 ATAGCCCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 319

RESULT 7  
CC819240 414 bp DNA linear GSS 17-JUL-2003  
LOCUS CC819240  
DEFINITION 100005D19R Oxytricha plasmid U0000 library Sterkiella  
histriomuscorum genomic clone U0000005D19 R, genomic survey  
Sequence.  
ACCESSION CC819240  
VERSION CC819240.1 GI:32899308  
KEYWORDS GSS.

```

SOURCE          Sterkiella histriomuscorum (Oxytricha trifallax)
ORGANISM        Sterkiella histriomuscorum
REFERENCE       Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
AUTHORS         Stichotrichidae; Oxytrichidae; Sterkiella.
TITLE           1 (bases 1 to 414)
JOURNAL         Dunn, D., Doak, T., Herrick, G. and Weiss, R.
COMMENT         Paired end reads from plasmid inserts of Oxytricha trifallax
                macronuclear chromosomes
                Unpublished (2003)
CONTACT        Contact: Robert B. Weiss
                University of Utah Genome Center
                University of Utah
                Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
                84112, USA
                Tel: 801 585 5606
                Fax: 801 585 7177
                Email: ddunn@genetics.utah.edu
                Plate: 0005 row: D column: 19
                Seq primer: CACACAGGAACACGCTAAGACC
                Class: plasmid ends
                High quality sequence stop: 414.
                Location/Qualifiers
                1..414
                /organism="Sterkiella histriomuscorum"
                /mol_type="genomic DNA"
                /db_xref="taxon:94289"
                /clone="UUGC100005D19"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /clone_1ib="Oxytricha plasmid UUGC10 library"
                /note="Vector: pMD42nv; Purified macronuclear chromosomal
                DNA from Oxytricha trifallax was blunt end-repaired with
                T4 DNA polymerase and T4 polynucleotide kinase. Adaptor
                oligonucleotides were ligated to the blunt ends in high
                molar excess. Vector DNA was prepared from a derivative of
                pMD42 [gi|4732114|gb|AF129072.1], a copy-number inducible
                derivative of plasmid R1. The vector was ligated with
                adaptors complementary to the insert adaptors and
                purified. The sheared, adaptor mouse DNA was annealed to
                adaptor vector DNA, and transformed into
                chemically-competent E. Coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
ORIGIN
Query Match      8.7%; Score 224.8; DB 29; Length 414;
Best Local Similarity 99.1%; Pred. No. 5.1e-58;
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2351 TACGCTGATGCCGATAGTTAAGCAGGCCCGCAGACCCGCCAACAACCCGCTGACGCGCC 2410
DB 87 TGCTGTGATGCCGATAGTTAAGCAGGCCCGCAGACCCGCCAACAACCCGCTGACGCGCC 146
QY 2411 TGACGGGCTGTCTCTCTCCCGCATCCGCTTACACAGACGTGTACCGTCTCCGGAGAC 2470
DB 147 TGACGGGCTGTCTCTCTCCCGCATCCGCTTACACAGACGTGTACCGTCTCCGGAGAC 206
QY 2471 TGCATGTGTACAGAGTTTTCACCGTCATCACCGAAGCGCGCAGACGAAAGGCTCTG 2530
DB 207 TGCATGTGTACAGAGTTTTCACCGTCATCACCGAAGCGCGCAGACGAAAGGCTCTG 266
QY 2531 ATAGCGCTATTTTATAGTTATGTGATGATGATTAATATGTTTCTTAG 2578
DB 267 ATAGCGCTATTTTATAGTTATGTGATGATTAATATGTTTCTTAG 314
RESULT 8
CC816935 464 bp DNA linear GSS 17-JUL-2003
LOCUS      CC816935
DEFINITION 100002C08R Oxytricha plasmid UUGC10 library Sterkiella
            histriomuscorum genomic clone UUGC100002C08 R, genomic survey
            sequence.
ACCESSION  CC816935
VERSION    CC816935.1 GI:32896222
KEYWORDS   GSS.

```

SOURCE	Sterkiella histriomuscorum (Oxytricha trifallax)		
ORGANISM	Sterkiella histriomuscorum		
REFERENCE	Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia; Stichotrichida; Oxytrichidae; Sterkiella.		
AUTHORS	Dunn, D., Doek, T., Herrick, G. and Weiss, R.		
TITLE	Paired end reads from plasmid inserts of Oxytricha trifallax macronuclear chromosomes		
JOURNAL	Unpublished (2003)		
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Plate: 0002 row: C column: 08 Seq primer: CACACGAGAAAACGCTATGACC Class: plasmid ends High quality sequence stop: 464. Location/Qualifiers		
FEATURES	1. 464 /organism="Sterkiella histriomuscorum" /mol_type="genomic DNA" /db_xref="taxon:94289" /clone="U06C100002C08" /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-" /clone_1bp="Oxytricha plasmid U06C10 library" /note="Vector: pMD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-ligated with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (G14732114[GB AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."		
ORIGIN	Query Match 8.7%; Score 224.8; DB 29; Length 464; Best local similarity 99.1%; Pred. No. 5,4e-58; Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY	2251	TACGCTGATGCCGATAGTTAAGCCAGCCCGCGACACCCGCCACACCCGCGCTGACGCGCC	2410
DB	180	TGCTCTGATGCCGATAGTTAAGCCAGCCCGCGACACCCGCCACACCCGCGCTGACGCGCC	239
QY	2411	TGACGGCGCTGTGCTCCCGCGCATCGGCTTAACAGACAGCTGTGACCGCTCCGGGAGC	2470
DB	240	TGACGGCGCTGTGCTCCCGCGCATCGGCTTAACAGACAGCTGTGACCGCTCCGGGAGC	229
QY	2471	TGCATGTGTCAAGAGTTTCAACCGCTCATCAACGGAACGCGGAGACGAAAGGCGCTCGT	2520
DB	300	TGCATGTGTCAAGAGTTTCAACCGCTCATCAACGGAACGCGGAGACGAAAGGCGCTCGT	359
QY	2531	ATAGCGCTATTTTATAGCTTAAGTCATGATATATATAGTTGTTGTTAG	2578
DB	360	ATAGCGCTATTTTATAGCTTAAGTCATGATATATATAGTTTCTTAG	407
RESULT 9	CC817727 471 bp DNA linear GSS 17-JUL-2003		
LOCUS	CC817727		
DEFINITION	100003007R Oxytricha plasmid U06C10 library Sterkiella histriomuscorum genomic clone U06C100003007 R, genomic survey sequence.		
ACCESSION	CC817727		
VERSION	CC817727.1 GI:32897014		
KEYWORDS	GSS.		



SOURCE  
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)  
REFERENCE Sterkiella histriomuscorum  
AUTHORS Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;  
TITLE Stichotrichida; Oxytrichidae; Sterkiella.  
1 (bases 1 to 471)  
Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
Paired end reads from plasmid inserts of Oxytricha trifallax  
macronuclear chromosomes  
JOURNAL Unpublished (2003)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0003 row: 0 column: 07  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 471.  
Location/Qualifiers

## FEATURES

## source

1..471  
/organism="Sterkiella histriomuscorum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:94289"  
/clone="UUC100006J13"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (g114732114|g114732114|g114732114), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 471;  
Best Local Similarity 99.1%; Pred. No. 5.5e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TACGCTGATGCGCATGTTAAGCCAGCCCGGCAACCCCGTGAAGCGCC 2410  
DB 91 TGCTGTATGCGCATGTTAAGCCAGCCCGGCAACCCCGTGAAGCGCC 150  
QY 2411 TGACGGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGTCCGGAGC 2470  
DB 151 TGACGGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGTCCGGAGC 210  
QY 2471 TGCATGTGTGAGAGGTTTTCACCGTCATCACCGAAACCGCGAGACGAAAGGGCTCTGT 2530  
DB 211 TGCATGTGTGAGAGGTTTTCACCGTCATCACCGAAACCGCGAGACGAAAGGGCTCTGT 270  
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATTAATGTTCTTAG 2578  
DB 271 ATAGCGCTATTTTATAGTTAATGTCATGATTAATGTTCTTAG 318

RESULT 10  
CC819923 491 bp DNA linear GSS 17-JUL-2003  
LOCUS CC819923  
DEFINITION 100006J13 Oxytricha plasmid UUC10 library Sterkiella  
histriomuscorum genomic clone UUC100006J13 R, genomic survey  
sequence.  
ACCESSION CC819923  
VERSION CC819923.1 GI:32900671  
KEYWORDS GSS.

SOURCE  
ORGANISM Sterkiella histriomuscorum (Oxytricha trifallax)  
REFERENCE Sterkiella histriomuscorum  
AUTHORS Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;  
TITLE Stichotrichida; Oxytrichidae; Sterkiella.  
1 (bases 1 to 491)  
Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
Paired end reads from plasmid inserts of Oxytricha trifallax  
macronuclear chromosomes  
JOURNAL Unpublished (2003)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0006 row: 0 column: 13  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 491.  
Location/Qualifiers

## FEATURES

## source

1..491  
/organism="Sterkiella histriomuscorum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:94289"  
/clone="UUC100006J13"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (g114732114|g114732114|g114732114), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 491;  
Best Local Similarity 99.1%; Pred. No. 5.6e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TACGCTGATGCGCATGTTAAGCCAGCCCGGCAACCCCGTGAAGCGCC 2410  
DB 85 TGCTGTATGCGCATGTTAAGCCAGCCCGGCAACCCCGTGAAGCGCC 144  
QY 2411 TGACGGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGTCCGGAGC 2470  
DB 145 TGACGGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAGCTGACCGTCCGGAGC 204  
QY 2471 TGCATGTGTGAGAGGTTTTCACCGTCATCACCGAAACCGCGAGACGAAAGGGCTCTGT 2530  
DB 205 TGCATGTGTGAGAGGTTTTCACCGTCATCACCGAAACCGCGAGACGAAAGGGCTCTGT 264  
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATTAATGTTCTTAG 2578  
DB 265 ATAGCGCTATTTTATAGTTAATGTCATGATTAATGTTCTTAG 312

RESULT 11  
AJ568083 507 bp mRNA linear EST 28-JUL-2003  
LOCUS AJ568083  
DEFINITION Antirrhinum majus library (Stueber K) Antirrhinum majus  
cDNA clone z935, mRNA sequence.  
ACCESSION AJ568083  
VERSION AJ568083.1 GI:33293972  
KEYWORDS EST  
SOURCE Antirrhinum majus (snapdragon)

ORGANISM  
Antirrhinum majus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamiales; Lamiales; Antirrhinaceae; Antirrhineae;  
Antirrhinum.  
1 (bases 1 to 507)  
Stueber, K. and Schwarz-Sommer, Z.  
Antirrhinum majus EST collection (Stueber, K. and Schwarz-Sommer, Z.)  
Unpublished (2003)  
CONTACT: Stueber K  
Molecular Plant Genetics  
MPI fuer Zuechtungsforchung  
Carl-von-Linne Weg 10, D-50829, Germany.  
Location/Qualifiers  
1. 507  
/organism="Antirrhinum majus"  
/mol\_type="mRNA"  
/db\_xref="taxon:4151"  
/clone="2835"  
/clone\_lib="Antirrhinum majus library (Stueber K)"

ORIGIN  
Query Match 8.7%; Score 224.8; DB 9; Length 507;  
Best Local Similarity 99.1%; Pred. No. 5.7e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCGCGATGTTAAGCCAGCCCGCAGACCCCGCAACACCCGCTGACGCGCC 2410  
DB 273 TGCTGTATGCGCGATGTTAAGCCAGCCCGCAGACCCCGCAACACCCGCTGACGCGCC 332  
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGTACCTCTCCGGAGC 2470  
DB 333 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGTACCTCTCCGGAGC 392  
QY 2471 TGATGTGTCAAGAGTTTTCACCGTCATCACCGAAGCGCGAGAGAGAGGCTCTGTG 2530  
DB 393 TGATGTGTCAAGAGTTTTCACCGTCATCACCGAAGCGCGAGAGAGGCTCTGTG 452  
QY 2531 ATACGCGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 2578  
DB 453 ATACGCGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 500

RESULT 12  
CC817752 515 bp DNA linear GSS 17-JUL-2003  
LOCUS CC817752  
DEFINITION 100003C16R Oxytricha plasmid UUGC10 library Sterkiella  
histriomuscorum genomic clone UUGC100003C16 R, genomic survey  
sequence.  
ACCESSION CC817752  
VERSION CC817752.1 GI:32897039  
KEYWORDS GSS.  
SOURCE Sterkiella histriomuscorum (Oxytricha trifallax)  
ORGANISM Sterkiella histriomuscorum  
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;  
Stichotrichida; Oxytrichidae; Sterkiella.  
1 (bases 1 to 515)  
Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
Paired end reads from plasmid inserts of Oxytricha trifallax  
macronuclear chromosomes  
Unpublished (2003)  
CONTACT: Robert B. Weiss  
University of Utah  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel.: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0003 row: C column: 16  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 515.

JOURNAL  
AUTHORS  
TITLE  
COMMENT

FEATURES  
source  
Location/Qualifiers  
1. 515  
/organism="Sterkiella histriomuscorum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:94289"  
/clone="UUGC100003C16"  
/db\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Oxytricha plasmid UUGC10 library"  
/note="Vector: PMD42nv; Purified macronuclear chromosomal  
DNA from Oxytricha trifallax was blunt end-repaired with  
T4 DNA polymerase and T4 polynucleotide kinase. Adaptor  
oligonucleotides were ligated to the blunt ends in high  
molar excess. Vector DNA was prepared from a derivative of  
pMD42 (G14732114|gb|AF129072.1), a copy-number inducible  
derivative of plasmid R1. The vector was ligated with  
adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. Coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

ORIGIN  
Query Match 8.7%; Score 224.8; DB 29; Length 515;  
Best Local Similarity 99.1%; Pred. No. 5.8e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTGATGCGCGATGTTAAGCCAGCCCGCAGACCCCGCAACACCCGCTGACGCGCC 2410  
DB 85 TGCTGTATGCGCGATGTTAAGCCAGCCCGCAGACCCCGCAACACCCGCTGACGCGCC 144  
QY 2411 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGTACCTCTCCGGAGC 2470  
DB 145 TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAGCTGTACCTCTCCGGAGC 204  
QY 2471 TGATGTGTCAAGAGTTTTCACCGTCATCACCGAAGCGCGAGAGAGAGGCTCTGTG 2530  
DB 205 TGATGTGTCAAGAGTTTTCACCGTCATCACCGAAGCGCGAGAGAGGCTCTGTG 264  
QY 2531 ATACGCGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 2578  
DB 265 ATACGCGCTATTTTATAGTTAATGTCATGATATATATAGTTCTTAG 312

RESULT 13  
CC817128 518 bp DNA linear GSS 17-JUL-2003  
LOCUS CC817128  
DEFINITION 100002D21R Oxytricha plasmid UUGC10 library Sterkiella  
histriomuscorum genomic clone UUGC100002D21 R, genomic survey  
sequence.  
ACCESSION CC817128  
VERSION CC817128.1 GI:32896415  
KEYWORDS GSS.  
SOURCE Sterkiella histriomuscorum (Oxytricha trifallax)  
ORGANISM Sterkiella histriomuscorum  
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;  
Stichotrichida; Oxytrichidae; Sterkiella.  
1 (bases 1 to 518)  
Dunn, D., Doak, T., Herrick, G. and Weiss, R.  
Paired end reads from plasmid inserts of Oxytricha trifallax  
macronuclear chromosomes  
Unpublished (2003)  
CONTACT: Robert B. Weiss  
University of Utah  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel.: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0002 row: D column: 21  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 518.

JOURNAL  
AUTHORS  
TITLE  
COMMENT

FEATURES  
source

Location/Qualifiers  
1. 518  
/organism="Sterkiella histriomuscorum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:94289"  
/clone="UUGC10002A23"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_1lb="Oxytricha plasmid UUGC10 library"  
/note="Vector: PMD42nv. Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 [gi14732114|gb|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 519;  
Best Local Similarity 99.1%; Pred. No. 5.8e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2351 TAGCGTATGCGCATAGTTAAGCCAGCCCGGACACCCGCAACCCGCTGAGCGGCC 2410  
83 TGTCTGATGCGCATAGTTAAGCCAGCCCGGACACCCGCAACCCGCTGAGCGGCC 142  
2411 TGAAGGCTGTGCTGCTCCCGCATCCGCTTAACAGACAGCTGTGACCGTCCGGAGC 2470  
143 TGAAGGCTGTGCTGCTCCCGCATCCGCTTAACAGACAGCTGTGACCGTCCGGAGC 202  
2471 TGCATGTGTCAGAGTTTCAACCGTCATCAACCGGAGAGCGGAGGAGGAGGAGG 2530  
203 TGCATGTGTCAGAGTTTCAACCGTCATCAACCGGAGAGCGGAGGAGGAGGAGG 262  
2531 ATAGCGCTATTTTATAGTTAATGTCATGATATATATATATATATATATATATAT 2578  
263 ATAGCGCTATTTTATAGTTAATGTCATGATATATATATATATATATATATATAT 310

RESULT 14  
CC817112 519 bp DNA linear GSS 17-JUL-2003  
LOCUS 100002A23R Oxytricha plasmid UUGC10 library Sterkiella  
DEFINITION histriomuscorum genomic clone UUGC100002A23 R, genomic survey  
sequence.  
ACCESSION CC817112  
VERSION CC817112.1 GI:32896399  
KEYWORDS GSS.  
SOURCE Sterkiella histriomuscorum (Oxytricha trifallax)  
ORGANISM Sterkiella histriomuscorum  
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;  
Stichotrichida; Oxytrichidae; Sterkiella.  
1 (bases 1 to 519)  
Dunn,D., Doak,T., Herrick,G. and Weiss,R.  
Paired end reads from plasmid inserts of Oxytricha trifallax  
macronuclear chromosomes  
Unpublished (2003)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0002 row: A column: 23  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 519.

FEATURES  
source

Location/Qualifiers  
1. 519  
/organism="Sterkiella histriomuscorum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:94289"  
/clone="UUGC10002A23"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_1lb="Oxytricha plasmid UUGC10 library"  
/note="Vector: PMD42nv. Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 [gi14732114|gb|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 519;  
Best Local Similarity 99.1%; Pred. No. 5.8e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2351 TAGCGTATGCGCATAGTTAAGCCAGCCCGGACACCCGCAACCCGCTGAGCGGCC 2410  
225 TGTCTGATGCGCATAGTTAAGCCAGCCCGGACACCCGCAACCCGCTGAGCGGCC 294  
2411 TGAAGGCTGTGCTGCTCCCGCATCCGCTTAACAGACAGCTGTGACCGTCCGGAGC 2470  
225 TGAAGGCTGTGCTGCTCCCGCATCCGCTTAACAGACAGCTGTGACCGTCCGGAGC 354  
2471 TGCATGTGTCAGAGTTTCAACCGTCATCAACCGGAGAGCGGAGGAGGAGGAGG 2530  
355 TGCATGTGTCAGAGTTTCAACCGTCATCAACCGGAGAGCGGAGGAGGAGGAGG 414  
2531 ATAGCGCTATTTTATAGTTAATGTCATGATATATATATATATATATATATATAT 2578  
415 ATAGCGCTATTTTATAGTTAATGTCATGATATATATATATATATATATATATAT 462

RESULT 15  
CC817796 519 bp DNA linear GSS 17-JUL-2003  
LOCUS 100003K14R Oxytricha plasmid UUGC10 library Sterkiella  
DEFINITION histriomuscorum genomic clone UUGC10003K14 R, genomic survey  
sequence.  
ACCESSION CC817796  
VERSION CC817796.1 GI:32897083  
KEYWORDS GSS.  
SOURCE Sterkiella histriomuscorum (Oxytricha trifallax)  
ORGANISM Sterkiella histriomuscorum  
Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;  
Stichotrichida; Oxytrichidae; Sterkiella.  
1 (bases 1 to 519)  
Dunn,D., Doak,T., Herrick,G. and Weiss,R.  
Paired end reads from plasmid inserts of Oxytricha trifallax  
macronuclear chromosomes  
Unpublished (2003)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Plate: 0003 row: K column: 14  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 519.

FEATURES  
source

Location/Qualifiers  
1. 519  
/organism="Sterkiella histriomuscorum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:94289"  
/clone="UUGC10003X14"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42ny; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 8.7%; Score 224.8; DB 29; Length 519;  
Best Local Similarity 99.1%; Pred. No. 5.8e-58;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	2351	TACGCTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACCCGCTGACGCGCC	2410
DB	88	TGCTTGATGCCGATAGTTAAGCCAGCCCGCAACCCCGCAACCCGCTGACGCGCC	147
QY	2411	TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGC	2470
DB	148	TGACGGGCTTGTCTGCTCCCGGATCCGCTTACAGACAAAGCTGTGACCGTCTCCGGAGC	207
QY	2471	TGCATGTGCAAGAGTTTCAACCGTATGACGAAACGGCGAGACGAAAGGCGCTCGTG	2530
DB	208	TGCATGTGCAAGAGTTTCAACCGTATGACGAAACGGCGAGACGAAAGGCGCTCGTG	267
QY	2531	ATAAGCCTATTTTATAGGTATATGATGATATATATGTTCTTAG	2578
DB	268	ATAAGCCTATTTTATAGGTATATGATGATATATATGTTCTTAG	315

Search completed: April 30, 2004, 18:34:55  
Job time : 7322.21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:32 ; Search time 302.986 Seconds

(without alignments)  
4721.887 Million cell updates/sec

Title: US-10-603-260-1

Perfect score: 2578

Sequence: 1 agctgcacttatcagcca.....tgataataatggtttcttag 2578

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_NA:  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	224.8	8.7	1073	US-09-481-049-6	Sequence 6, Appl
2	224.8	8.7	2580	US-09-194-285-23	Sequence 23, Appl
3	224.8	8.7	2939	US-08-119-512-2	Sequence 2, Appl
4	224.8	8.7	2939	US-08-488-015B-2	Sequence 2, Appl
5	224.8	8.7	2939	US-08-814-412-17	Sequence 17, Appl
6	224.8	8.7	3104	US-07-415-307A-1	Sequence 1, Appl
7	224.8	8.7	3104	US-08-371-320-1	Sequence 1, Appl
8	224.8	8.7	3190	US-09-027-169-6	Sequence 6, Appl
9	224.8	8.7	3343	5453363-2	Parent No. 5453363
10	224.8	8.7	3796	US-09-470-661A-32	Sequence 32, Appl
11	224.8	8.7	3799	US-09-173-053-4	Sequence 4, Appl
12	224.8	8.7	3819	US-09-042-353-393	Sequence 393, App
C 13	224.8	8.7	3819	US-08-758-417A-243	Sequence 243, Appl
14	224.8	8.7	3875	US-09-039-982A-21	Sequence 21, Appl
15	224.8	8.7	3875	US-09-039-641-21	Sequence 21, Appl
16	224.8	8.7	3875	US-09-039-762A-21	Sequence 21, Appl
17	224.8	8.7	3875	US-09-042-492D-21	Sequence 21, Appl
18	224.8	8.7	3875	US-08-913-612A-21	Sequence 21, Appl
19	224.8	8.7	3878	US-09-039-982A-27	Sequence 27, Appl
20	224.8	8.7	3878	US-09-039-641-27	Sequence 27, Appl
21	224.8	8.7	3878	US-09-039-762A-27	Sequence 27, Appl
22	224.8	8.7	3878	US-09-042-492D-27	Sequence 27, Appl
23	224.8	8.7	3878	US-08-913-612A-27	Sequence 27, Appl
24	224.8	8.7	3881	US-09-042-353-369	Sequence 369, App
C 25	224.8	8.7	3881	US-08-758-417A-217	Sequence 217, App
26	224.8	8.7	3883	US-09-039-982A-30	Sequence 30, Appl
27	224.8	8.7	3883	US-09-039-641-30	Sequence 30, Appl

28	224.8	8.7	3883	US-09-039-762A-30	Sequence 30, Appl
29	224.8	8.7	3883	US-09-042-492D-30	Sequence 30, Appl
30	224.8	8.7	3883	US-08-813-612A-30	Sequence 30, Appl
31	224.8	8.7	3908	US-09-039-982A-24	Sequence 24, Appl
32	224.8	8.7	3908	US-09-039-641-24	Sequence 24, Appl
33	224.8	8.7	3908	US-09-039-762A-24	Sequence 24, Appl
34	224.8	8.7	3908	US-09-042-492D-24	Sequence 24, Appl
35	224.8	8.7	3908	US-08-913-612A-24	Sequence 24, Appl
36	224.8	8.7	3984	US-08-040-753-1	Sequence 1, Appl
37	224.8	8.7	4016	US-09-173-053-3	Sequence 3, Appl
38	224.8	8.7	4045	US-08-464-700-54	Sequence 54, Appl
39	224.8	8.7	4054	US-09-098-287A-9	Sequence 9, Appl
40	224.8	8.7	4054	US-09-445-649-9	Sequence 9, Appl
C 41	224.8	8.7	4118	US-09-068-821-17	Sequence 17, Appl
C 42	224.8	8.7	4118	US-09-068-821-18	Sequence 18, Appl
C 43	224.8	8.7	4161	US-09-185-244-8	Sequence 8, Appl
44	224.8	8.7	4161	US-09-471-913-1	Sequence 1, Appl
C 45	224.8	8.7	4273	US-08-795-430-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-481-049-6/c  
Sequence 6, Application US/09481049  
Patent No. 6395485  
GENERAL INFORMATION:  
APPLICANT: DE BEUCKELER, Marc  
TITLE OF INVENTION: METHODS FOR IDENTIFYING ELITE EVENT GAT-ZM1 IN  
FILE REFERENCE: 514412-2025  
CURRENT APPLICATION NUMBER: US/09/481,049  
CURRENT FILING DATE: 2000-01-11  
NUMBER OF SEQ. ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 1073  
TYPE: DNA  
ORGANISM: Zea mays  
US-09-481-049-6

Query Match 8.7%; Score 224.8; DB 4; Length 1073;  
Best Local Similarity 99.1%; Pred. No. 7e-60; 2; Indels 0; Gaps 0;  
Matches 226; Conservative 0; Mismatches 2

Cy 2351 TAGCTGATGCCGATAGTTAAGCCAGCCCGCAGCAACCCGCTGACGGCC 2410  
Db 709 TGCTGTGATGCCGATAGTTAAGCCAGCCCGCAGCAACCCGCTGACGGCC 650  
Cy 2411 TGACGGGCTGTGCTCCCGGATCCGCTTACAGCAAGCTGTGACCGTCCGGGAGC 2470  
Db 649 TGACGGGCTGTGCTCCCGGATCCGCTTACAGCAAGCTGTGACCGTCCGGGAGC 590  
Cy 2471 TGATGTGTGACAGGTTTCAACCGTATCATCCGCAACCGCGGAGAGAGGCGCTCGTG 2530  
Db 589 TGATGTGTGACAGGTTTCAACCGTATCATCCGCAACCGCGGAGAGAGGCGCTCGTG 530  
Cy 2531 ATAGGCTATTTTATAGGTTAATGTCATGATTAATATGTTCTTAG 2578  
Db 529 ATAGGCTATTTTATAGGTTAATGTCATGATTAATATGTTCTTAG 482

RESULT 2  
US-09-194-285-23  
Sequence 23, Application US/09194285  
Patent No. 6355479  
GENERAL INFORMATION:  
APPLICANT: Webd, Susan R.  
APPLICANT: Windqvist, Oia  
APPLICANT: Karlsson, Lars  
APPLICANT: Jackson, Michael R.  
APPLICANT: Peterson, Per A.

TITLE OF INVENTION: MEC Class II Antigen Presenting Systems  
TITLE OF INVENTION: and Methods for Activating CD4+ T Cells  
FILE REFERENCE: TSRI 536.1  
CURRENT APPLICATION NUMBER: US/09/194,285  
PRIOR FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: PCT/US97/08697  
PRIOR FILING DATE: 1997-05-22  
PRIOR APPLICATION NUMBER: US 60/018,175  
PRIOR FILING DATE: 1996-05-23  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23  
LENGTH: 2580  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: engineered recombinant  
US-09-194-285-23

Query Match 8.7%; Score 224.8; DB 4; Length 2580;  
Best Local Similarity 99.1%; Pred. No. 1.2e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTATGCCGATAGTTAAGCAGCCCGCAGACCCCGCAACCCGCTGACGCGCC 2410  
DB 2126 TGCTCTATGCCGATAGTTAAGCAGCCCGCAGACCCCGCAACCCGCTGACGCGCC 2185  
QY 2411 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTACCGCTTCCGGAGC 2470  
DB 2186 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTACCGCTTCCGGAGC 2245  
QY 2471 TGATGTGTGAGAGGTTTACCGCTATCAGCCGATCAGAAAGGCGGCGCTCGTG 2530  
DB 2246 TGCTGTGTGAGAGGTTTACCGCTATCAGCCGATCAGAAAGGCGGCGGAGGCGCTCGTG 2305  
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATGATATATATATGTTCTTAG 2578  
DB 2306 ATAGCGCTATTTTATAGTTAATGTCATGATGATATATATGTTCTTAG 2353

RESULT 3  
US-08-119-512-2  
Sequence 2, Application US/08119512  
Patent No. 5498531  
GENERAL INFORMATION:  
APPLICANT: Jarrell, Kevin A.  
TITLE OF INVENTION: INTRON MEDIATED RECOMBINANT TECHNIQUES  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/119,512  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: HUI-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2939 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: both  
MOLECULE TYPE: other nucleic acid  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2448..2657  
OTHER INFORMATION: /product= "b-globin exon 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2667..2814  
OTHER INFORMATION: /product= "b-globin exon 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2815..2890  
OTHER INFORMATION: /product= "intron sequence"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2390..2447  
OTHER INFORMATION: /product= "intron sequence"  
US-08-119-512-2

Query Match 8.7%; Score 224.8; DB 1; Length 2939;  
Best Local Similarity 99.1%; Pred. No. 1.3e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TACGCTATGCCGATAGTTAAGCAGCCCGCAGACCCCGCAACCCGCTGACGCGCC 2410  
DB 105 TGCTCTATGCCGATAGTTAAGCAGCCCGCAGACCCCGCAACCCGCTGACGCGCC 164  
QY 2411 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTACCGCTTCCGGAGC 2470  
DB 165 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGACAGCTGTACCGCTTCCGGAGC 224  
QY 2471 TGATGTGTGAGAGGTTTACCGCTATCAGCCGATCAGAAAGGCGGCGCTCGTG 2530  
DB 225 TGATGTGTGAGAGGTTTACCGCTATCAGCCGATCAGAAAGGCGGCGGAGGCGCTCGTG 284  
QY 2531 ATAGCGCTATTTTATAGTTAATGTCATGATGATATATATGTTCTTAG 2578  
DB 285 ATAGCGCTATTTTATAGTTAATGTCATGATGATATATATGTTCTTAG 332

RESULT 4  
US-08-488-015B-2  
Sequence 2, Application US/08488015B  
Patent No. 5780272  
GENERAL INFORMATION:  
APPLICANT: Jarrell, Kevin A.  
TITLE OF INVENTION: INTRON-MEDIATED RECOMBINANT TECHNIQUES  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley, Hoag & Eliot  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,015B  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: HUV-008.02

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 832-1000  
TELEFAX: (617) 832-7000  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2939 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: both  
MOLECULE TYPE: other nucleic acid  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2448..2657  
OTHER INFORMATION: /product= "D-globin exon 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2667..2814  
OTHER INFORMATION: /product= "D-globin exon 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2815..2890  
OTHER INFORMATION: /product= "intron sequence"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2390..2447  
OTHER INFORMATION: /product= "intron sequence"  
US-08-488-015B-2  
Query Match 8.7%; Score 224.8; DB 1; Length 2939;  
Best Local Similarity 99.1%; Pred. No. 1.3e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TACGCTGATGCGCATAGTTAAGTAAAGCCGCGACACCCGCGTGAAGCGGCC 2410  
DB 105 TCGCTGATGCGCATAGTTAAGTAAAGCCGCGACACCCGCGTGAAGCGGCC 164  
QY 2411 TGACGGGCTGTGCTGCTCCGCGCATCCGCTTACAGACAGCTGTGACCGTCTCGGGAGC 2470  
DB 165 TGACGGGCTGTGCTGCTCCGCGCATCCGCTTACAGACAGCTGTGACCGTCTCGGGAGC 224  
QY 2471 TGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTCGTG 2530  
DB 225 TGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTCGTG 284  
QY 2531 ATAGCGCTATTTTATAGTTATGTCATGATTAATATGTTCTTAG 2578  
DB 285 ATAGCGCTATTTTATAGTTATGTCATGATTAATATGTTCTTAG 332  
RESULT 5  
US-08-814-412-17  
Sequence 17, Application US/08814412  
Patent No. 6150141  
GENERAL INFORMATION:  
APPLICANT: Jarrell Ph.D., Kevin A.  
TITLE OF INVENTION: Intron-Mediated Techniques and Reagents  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chateau, Hall & Stewart  
STREET: 53 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/814,412  
FILING DATE: 11-MAR-1997  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Jarrell Ph.D., Brenda H.  
REGISTRATION NUMBER: 39,223  
REFERENCE/DOCKET NUMBER: 0079571-0040  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 248 4000  
TELEFAX: 617 248 4000  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2939 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "plasmid DNA"  
IMMEDIATE SOURCE:  
CLONE: BGINV  
US-08-814-412-17  
Query Match 8.7%; Score 224.8; DB 3; Length 2939;  
Best Local Similarity 99.1%; Pred. No. 1.3e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TACGCTGATGCGCATAGTTAAGTAAAGCCGCGACACCCGCGTGAAGCGGCC 2410  
DB 105 TCGCTGATGCGCATAGTTAAGTAAAGCCGCGACACCCGCGTGAAGCGGCC 164  
QY 2411 TGACGGGCTGTGCTGCTCCGCGCATCCGCTTACAGACAGCTGTGACCGTCTCGGGAGC 2470  
DB 165 TGACGGGCTGTGCTGCTCCGCGCATCCGCTTACAGACAGCTGTGACCGTCTCGGGAGC 224  
QY 2471 TGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTCGTG 2530  
DB 225 TGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAGCGCGAGACGAAAGGCGCTCGTG 284  
QY 2531 ATAGCGCTATTTTATAGTTATGTCATGATTAATATGTTCTTAG 2578  
DB 285 ATAGCGCTATTTTATAGTTATGTCATGATTAATATGTTCTTAG 332  
RESULT 6  
US-07-415-307A-1/C  
Sequence 1, Application US/07415307A  
Patent No. 5344757  
GENERAL INFORMATION:  
APPLICANT: Holte, Hans Joachim  
APPLICANT: Seib, Rudolf  
APPLICANT: Schmitz, Gudrun  
APPLICANT: Kessler, Hans  
APPLICANT: Kessler, Christoph  
APPLICANT: Matthes, Ralf  
TITLE OF INVENTION: Process for the Detection of Nucleic Acids  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felte & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC/DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/415,307A  
FILING DATE: 19900109  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP89/0026  
FILING DATE: 12-Jan-1989  
APPLICATION NUMBER: DE 38 00 642.1



FILING DATE: 12-Jan-1988  
 APPLICATION NUMBER: DE 38 13 278.8  
 FILING DATE: 20-NOV-1988  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hanson, No. 5344757man D.  
 REGISTRATION NUMBER: 30,946  
 REFERENCE/DOCKET NUMBER: BOER 798  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-668-8200  
 TELEFAX: 212-838-3884  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3104 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-07-415-307A-1

	Query Match	8.7%	Score 224.8	DB 1	length 3104	Best Local Similarity 99.1%	Pred. No. 1,4e-59	Matches 226	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	2351	TACGTGATGCCGCATAGTTAAAGCCAGCCCGGACACCCCGCCGATGACCGGCC	2411	TACGTGATGCCGCATAGTTAAAGCCAGCCCGGACACCCCGCCGATGACCGGCC	2411							
Db	2558	TGCTGTGATGCCGCATAGTTAAAGCCAGCCCGGACACCCCGCCGATGACCGGCC	2498	TGCTGTGATGCCGCATAGTTAAAGCCAGCCCGGACACCCCGCCGATGACCGGCC	2498							
QY	2411	TGACGGGGTTGTCTGCTCCCGGCATCCGGCTTACACAAAGCGTGTCTCCGGGAGC	2470	TGACGGGGTTGTCTGCTCCCGGCATCCGGCTTACACAAAGCGTGTCTCCGGGAGC	2470							
Db	2498	TGACGGGGTTGTCTGCTCCCGGCATCCGGCTTACACAAAGCGTGTCTCCGGGAGC	2438	TGACGGGGTTGTCTGCTCCCGGCATCCGGCTTACACAAAGCGTGTCTCCGGGAGC	2438							
QY	2471	TGCATGTGTCAAGGTTTTCAACCGTCATCACCGAAACGGCGAGACGAAAGGGCCTCGTG	2530	TGCATGTGTCAAGGTTTTCAACCGTCATCACCGAAACGGCGAGACGAAAGGGCCTCGTG	2530							
Db	2438	TGCATGTGTCAAGGTTTTCAACCGTCATCACCGAAACGGCGAGACGAAAGGGCCTCGTG	2379	TGCATGTGTCAAGGTTTTCAACCGTCATCACCGAAACGGCGAGACGAAAGGGCCTCGTG	2379							
QY	2531	ATAGCGCCATTTTATAGTTAATGTCAATGATTAATAATGGTTTCTTAG	2578	ATAGCGCCATTTTATAGTTAATGTCAATGATTAATAATGGTTTCTTAG	2578							
Db	2378	ATAGCGCCATTTTATAGTTAATGTCAATGATTAATAATGGTTTCTTAG	2331	ATAGCGCCATTTTATAGTTAATGTCAATGATTAATAATGGTTTCTTAG	2331							

? APPLICATION NUMBER: DE 38 00 642.1  
 ? FILING DATE: 12-Jan-1988  
 ? APPLICATION NUMBER: DE 38 13 278.6  
 ? FILING DATE: 20-SEP-1988  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Hanson, No. 5702888man D.  
 ? REGISTRATION NUMBER: 30,946  
 ? REFERENCE/DOCKET NUMBER: BOER 798  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: 212-688-9200  
 ? TELEFAX: 212-83-8384  
 ? INFORMATION FOR SEQ ID NO: 1:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 3104 base pairs  
 ? TYPE: nucleic acid  
 ? STRANDEDNESS: single  
 ? TOPOLOGY: linear  
 ?  
 JS-08-371-320-1

	Query Match	Best Local Similarity	8.7% ; Score 224.8 ; DB 1 ; Length 3104 ;
	Matches	226 ; Conservative	0 ; Mismatches 2 ; Indels 0 ; Gaps 0 ;
QY	2351	TACGCTGATGCCGCAAGTTAAAGCCAGCCCCCGACACCCCGCCACACCCGCTGACGCGCC	2410
Db	2558	TGCTCTGATGCCGATGTTAAAGCAGCCCCGACACCCGCCAACACCCGCTGAGCGCGCC	2459
QY	2411	TGACGGGCTGTGTGCGACCCGGGCAATCCGCTTACAGACAGCTGACCGTTCCTCCGGAGC	2470
Db	2498	TGACGGGCTGTGTCTCCCGGCATCCGCTTACAGACAGCTGTGACCGTTCCTCCGGAGC	2439
QY	2471	TGCATGTGTGAGAGGTTTCAACCGTCATCAACGAAACGCGCAGACGAAAGGCGCTGTG	2550
Db	2438	TGCATGTGTGAGAGGTTTTCACCGTCATCAACGAAACGCGCAGACGAAAGGCGCTGTG	2379
QY	2531	ATAGCGCTATTTTAAATAGTTAATGTAATGTAATATATATGTTCTTTG	2578
Db	2378	ATAGCGCTATTTTAAATAGTTAATGTAATGTAATATATATGTTCTTTG	2331

SEQUENCE CHARACTERISTICS:  
LENGTH: 3190 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: circular  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "pRM2 (target plasmid)"  
US-09-027-169-6

Query Match 8.7%; Score 224.8; DB 4; Length 3190;  
Best Local Similarity 99.1%; Pred. No. 1.4e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTGAATGCGGATGTTAAGTAAAGCCGCGGACCGCCGCAACCCCGCTGACCGGCC 2410  
DB 1026 TGCTGTGATGCCGCAATGTTAAGCCGCGGACCGCCGCAACCCCGCTGACCGGCC 1085  
QY 2411 TGACGGGCTTGTCTGCTCCCGGCAATCGCTTACAGCAAGCTGTGACCGTCTCCGGAGC 2470  
DB 1086 TGACGGGCTTGTCTGCTCCCGGCAATCGCTTACAGCAAGCTGTGACCGTCTCCGGAGC 1145  
QY 2471 TGATGTGTCAAGAGTTTTCACCGTATACCGGAAACGGCGGAGAGGAGGCTCTGTG 2530  
DB 1146 TGATGTGTCAAGAGTTTTCACCGTATACCGGAAACGGCGGAGAGGAGGCTCTGTG 1205  
QY 2531 ATACGCTATTTTATAGTTAATGATGATATATATATATGTTTCTTAG 2578  
DB 1206 ATACGCTATTTTATAGTTAATGATGATATATATATATGTTTCTTAG 1253

RESULT 9  
545363-2  
PATENT NO. 545363  
APPLICANT: RUDOLPH, RAINER; FISCHER, STEPHAN; MATTES, RALF  
TITLE OF INVENTION: PROCESS FOR THE ACTIVATION OF T-PA OR  
ING AFTER GENETIC EXPRESSION IN PROKARYOTES  
NUMBER OF SEQUENCES: 4  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/206,044  
FILING DATE: 02-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 942,370  
FILING DATE: 09-SEP-1992  
APPLICATION NUMBER: 498,500  
FILING DATE: 23-MAR-1990  
APPLICATION NUMBER: 76,207  
FILING DATE: 23-OCT-1986  
SEQ ID NO: 2  
LENGTH: 3343  
545363-2

Query Match 8.7%; Score 224.8; DB 6; Length 3343;  
Best Local Similarity 99.1%; Pred. No. 1.4e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTGAATGCGGATGTTAAGTAAAGCCGCGGACCGCCGCAACCCCGCTGACCGGCC 2410  
DB 1179 TGCTGTGATGCCGCAATGTTAAGCCGCGGACCGCCGCAACCCCGCTGACCGGCC 1238  
QY 2411 TGACGGGCTTGTCTGCTCCCGGCAATCGCTTACAGCAAGCTGTGACCGTCTCCGGAGC 2470  
DB 1239 TGACGGGCTTGTCTGCTCCCGGCAATCGCTTACAGCAAGCTGTGACCGTCTCCGGAGC 1298  
QY 2471 TGATGTGTCAAGAGTTTTCACCGTATACCGGAAACGGCGGAGAGGAGGCTCTGTG 2530  
DB 1299 TGATGTGTCAAGAGTTTTCACCGTATACCGGAAACGGCGGAGAGGAGGCTCTGTG 1358  
QY 2531 ATACGCTATTTTATAGTTAATGATGATATATATATGTTTCTTAG 2578  
DB 1359 ATACGCTATTTTATAGTTAATGATGATATATATATGTTTCTTAG 1406

RESULT 10

US-09-470-661A-32/c  
Sequence 32, Application US/09470661A  
Patent No. 6500662  
GENERAL INFORMATION:  
APPLICANT: Pfizer Products Inc.  
TITLE OF INVENTION: AN INFECTIOUS CDNA CLONE OF NORTH AMERICAN PORCINE  
REPRODUCTIVE AND RESPIRATORY SYNDROME (PRRS) VIRUS AND  
TITLE OF INVENTION: USES THEREOF  
FILE REFERENCE: PC10278A  
CURRENT APPLICATION NUMBER: US/09/470,661A  
CURRENT FILING DATE: 1999-12-22  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 32  
LENGTH: 3796  
TYPE: DNA  
ORGANISM: Plasmid  
FEATURE:  
OTHER INFORMATION: Description of Plasmid: pCMV-MC1  
US-09-470-661A-32

Query Match 8.7%; Score 224.8; DB 4; Length 3796;  
Best Local Similarity 99.1%; Pred. No. 1.5e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGCGTGAATGCGGATGTTAAGTAAAGCCGCGGACCGCCGCAACCCCGCTGACCGGCC 2410  
DB 3564 TGCTGTGATGCCGCAATGTTAAGCCGCGGACCGCCGCAACCCCGCTGACCGGCC 3505  
QY 2411 TGACGGGCTTGTCTGCTCCCGGCAATCGCTTACAGCAAGCTGTGACCGTCTCCGGAGC 2470  
DB 3504 TGACGGGCTTGTCTGCTCCCGGCAATCGCTTACAGCAAGCTGTGACCGTCTCCGGAGC 3445  
QY 2471 TGATGTGTCAAGAGTTTTCACCGTATACCGGAAACGGCGGAGAGGAGGCTCTGTG 2530  
DB 3444 TGATGTGTCAAGAGTTTTCACCGTATACCGGAAACGGCGGAGAGGAGGCTCTGTG 3385  
QY 2531 ATACGCTATTTTATAGTTAATGATGATATATATATATGTTTCTTAG 2578  
DB 3384 ATACGCTATTTTATAGTTAATGATGATATATATATATGTTTCTTAG 3337

RESULT 11  
US-09-173-053-4/c  
Sequence 4, Application US/09173053  
Patent No. 6451769  
GENERAL INFORMATION:  
APPLICANT: HIEBNER, Robert C.  
APPLICANT: NORMAN, Jon A.  
APPLICANT: LIANG, Xiaowu  
APPLICANT: CARNER, Kristin R.  
APPLICANT: BARBOUR, Alan G.  
APPLICANT: LUXE, Catherine J.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ADMINISTERING BORRELIA DNA  
FILE REFERENCE: 454312-2440.1  
CURRENT APPLICATION NUMBER: US/09/173,053  
CURRENT FILING DATE: 1998-10-15  
PRIOR APPLICATION NUMBER: 08/663,998  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 3799  
TYPE: DNA  
ORGANISM: Borrelia burgdorferi  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1118)..(1125)  
OTHER INFORMATION: N stands for A or G or C or T  
NAME/KEY: misc feature  
LOCATION: (1130)..(1137)  
OTHER INFORMATION: N stands for A or G or C or T  
US-09-173-053-4

Query	Match	Best Local Similarity	9.7%;	Score 224.8;	DB 4;	Length 3799;
Matches 226;	Conservative	99.1%;	Pred. No. 1.5e-55;	0;	Mismatches 2;	Indels 0;
						Gaps 0;
QY	2351	TACGCTGATGCCGACATAGTTAAAGCCAGCCCGGACACCCGCGCAACACCCGCTGACGGGCC	2410			
Db	3566	TGCTCTGATGCCGACATAGTTAAAGCCAGCCCGACACCCGCGCAACACCCGCTGACGGGCC	3507			
QY	2411	TGACGGGCTGTCTGTCTCCCGGACATCGGCTTACAGACAACTGTGACCGCTCCCGGAAAC	2477			
Db	3506	TGACGGGCTGTGTGTCTCCCGGACATCGGCTTACAGACAACTGTGACCGCTCCCGGAAAC	3444			
QY	2471	TGCATGTGCAGAGGTTTTCAACCGTCATACCGAAACGCGCGACAGAAAGGCGCTGTG	2530			
Db	3446	TGCATGTGCAGAGGTTTTCAACCGTCATACCGAAACGCGCGAGCAAAAGGCGCTGTG	3387			
QY	2531	ATACGGCTATTTTAAAGTTAAATCTCATAGTAATTAATAGTTCTTAG	2578			
Db	3386	ATACGGCTATTTTAAAGTTAAATCTCATAGTAATTAATAGTTCTTAG	3339			

RESULT 12  
 US-09-042-353-393/C  
 Sequence 393, Application US/09042353  
 Patent No. 6253458  
 GENERAL INFORMATION:  
 APPLICANT: Lomborg, Nils  
 APPLICANT: Kay, Robert M.  
 TITLE OF INVENTION: Transgenic No.-6255458-Human Animals for  
 TITLE OF INVENTION: Producing Heterologous Antibodies  
 NUMBER OF SEQUENCES: 421  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-3834  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/042,353  
 FILING DATE: 13-MAR-1998  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/810,279  
 FILING DATE: 17-DEC-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/853,408  
 FILING DATE: 18-MAR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/904,068  
 FILING DATE: 23-JUN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/990,860  
 FILING DATE: 16-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/053,131  
 FILING DATE: 26-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/096,762  
 FILING DATE: 22-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/155,301  
 FILING DATE: 18-NOV-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/161,739  
 FILING DATE: 03-DEC-1993  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/165,699  
 FILING DATE: 10-DEC-1993  
 PRIOR APPLICATION DATA: US 08/209,741  
 APPLICATION NUMBER: US 08/165,699  
 FILING DATE: 09-MAR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/352,322  
 FILING DATE: 07-DEC-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/544,404  
 FILING DATE: 10-OCT-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/728,463  
 FILING DATE: 10-OCT-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/US96/16433  
 FILING DATE: 10-OCT-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/758,417  
 FILING DATE: 02-DEC-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/US97/21803  
 FILING DATE: 01-DEC-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Apple, Randolph T.  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 014643-009040US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 393:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3819 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA

	Query Match	Similarity	8.7%	Score	22.4.8	DB	3	Length	381.9	
	Match	Local	Similarity	99.18				Pred.	1.5e-59	
	Matches	226	Conservative	0	Mismatches	2	Indels	0	Gaps	0
QY	2351	TACGCGATGCGCGATGTTAAGCCAGACCCCGGACACCCCGCTGACGCGCC	2410							
Db	2165	TGCTGTATGCGCATGTAAAGCCACCCCGAACCCGCCAACCCGCTGACGCGCC	2106							
QY	2411	TGAGGGCTTGTCTGCTCCCGGATCCGGTTACAGAACTGTGACCGTCTCCGGAGC	2470							
Db	2105	TGAGGGCTTGTCTGCTCCCGGATCCGGTTACAGAACTGTGACCGTCTCCGGAGC	2046							
QY	2471	TGCATGTGTCAAGAGTTTCAACCGTCAACCGAAACCGCGGAGACGAAAGGCTCGTG	2530							
Db	2045	TGCATGTGTCAAGAGTTTCAACCGTCAACCGAAAGCGCGGAGACGAAAGGCTCGTG	1986							
QY	2531	ATAGCGTATTTTATAGTTAAGTAAAGTAAATATATAGTTTGTAG	2578							
Db	1985	ATAGCGCTATTTTATAGTTAAGTAAAGTAAATATATAGTTTGTAG	1938							

RESULT 13  
 US-08-758-417A-243/C  
 ; Sequence 243, Application US/08758417A  
 ; Patent No. 6300129  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lomborg, Nils  
 ; Kay, Robert M.  
 ; TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for  
 ; Producing Heterologous Antibodies  
 ; NUMBER OF SEQUENCES: 417  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Center, Eighth Floor  
 ;

CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/758,417A  
FILING DATE: 02-Dec-1996  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/728,463  
FILING DATE: 10-OCT-1996  
APPLICATION NUMBER: US 08/544,404  
FILING DATE: 10-OCT-1995  
APPLICATION NUMBER: US 08/352,322  
FILING DATE: 07-DEC-1994  
APPLICATION NUMBER: US 08/209,741  
FILING DATE: 09-MAR-1994  
APPLICATION NUMBER: US 08/165,699  
FILING DATE: 10-DEC-1993  
APPLICATION NUMBER: US 08/161,739  
FILING DATE: 03-DEC-1993  
APPLICATION NUMBER: US 08/155,301  
FILING DATE: 18-NOV-1993  
APPLICATION NUMBER: US 08/096,762  
FILING DATE: 22-JUL-1993  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Serafini, Andrew T.  
REGISTRATION NUMBER: 41,303  
REFERENCE/DOCKET NUMBER: 014643-009030US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 243:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3819 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 243:  
US-08-758-417A-243  
Query Match 8.7%; Score 224.8; DB 4; Length 3819;  
Best Local Similarity 99.1%; Pred. No. 1.5e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TAGCGTATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACACCCGCTGAGCGGCC 2410  
DB 2165 TGCTGTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACACCCGCTGAGCGGCC 2106  
QY 2411 TGACGGGCTTGCTGCTCCCGGATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470  
DB 2105 TGACGGGCTTGCTGCTCCCGGATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2046  
QY 2471 TGCAATGTCAAGAGTTTCAACCGTATACCGAAACCGCGAGACGAAAGGAGCTCTGTG 2530  
DB 2045 TGCAATGTCAAGAGTTTCAACCGTATACCGAAACCGCGAGACGAAAGGAGCTCTGTG 1986  
QY 2531 ATAGCCCTATTTTATAGATTATGTCATGATTAATATGTTCTTAG 2578  
DB 1985 ATAGCCCTATTTTATAGATTATGTCATGATTAATATGTTCTTAG 1938  
RESULT 14

US-09-039-982A-21  
Sequence 21, Application US/09039982A  
Patent No. 6225042  
GENERAL INFORMATION:  
APPLICANT: Cai, Zeling  
APPLICANT: Sprent, Jonathan  
APPLICANT: Brunmark, Anders  
APPLICANT: Jackson, Michael  
APPLICANT: Peterson, Per A  
TITLE OF INVENTION: ANTISENSE PRESENTING SYSTEM AND METHODS FOR ACTIVATION OF T-CELL  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Olsson & Hierrl, Ltd.  
STREET: 20 No. 6225042th Hacker Drive, Suite 3000  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/039,982A  
FILING DATE: 16-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Olsson, Arne M.  
REGISTRATION NUMBER: 30,203  
REFERENCE/DOCKET NUMBER: TSRI4710  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 580-1180  
TELEFAX: (312) 580-1189  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3875 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-039-982A-21  
Query Match 8.7%; Score 224.8; DB 3; Length 3875;  
Best Local Similarity 99.1%; Pred. No. 1.6e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TAGCGTATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACACCCGCTGAGCGGCC 2410  
DB 1482 TGCTGTGATGCCGATAGTTAAGCCAGCCCGACACCCCGCAACACCCGCTGAGCGGCC 1541  
QY 2411 TGACGGGCTTGCTGCTCCCGGATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 2470  
DB 1542 TGACGGGCTTGCTGCTCCCGGATCCGCTTACAGACAGCTGTGACCGTCTCCGGAGC 1601  
QY 2471 TGCAATGTCAAGAGTTTCAACCGTATACCGAAACCGCGAGACGAAAGGAGCTCTGTG 2530  
DB 1602 TGCAATGTCAAGAGTTTCAACCGTATACCGAAACCGCGAGACGAAAGGAGCTCTGTG 1661  
QY 2531 ATAGCCCTATTTTATAGATTATGTCATGATTAATATGTTCTTAG 2578  
DB 1662 ATAGCCCTATTTTATAGATTATGTCATGATTAATATGTTCTTAG 1709  
RESULT 15  
US-09-039-641-21  
Sequence 21, Application US/09039641  
Patent No. 6251627  
GENERAL INFORMATION:  
APPLICANT: Cai, Zeling  
APPLICANT: Sprent, Jonathan

APPLICANT: Brunmark, Anders  
APPLICANT: Jackson, Michael  
APPLICANT: Peterson, Per A  
TITLE OF INVENTION: ANTIGEN PRESENTING SYSTEM AND METHODS FOR  
TITLE OF INVENTION: ACTIVATION OF T-CELLS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Olson & Hierl, Ltd.  
STREET: 20 No. 6251627th Wacker Drive, Suite 3000  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/039,641  
FILING DATE: 8-MAR-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Olson, Arne M.  
REGISTRATION NUMBER: 30,203  
REFERENCE/DOCKET NUMBER: TSRI4710  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 580-1180  
TELEFAX: (312) 580-1189  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3875 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-039-641-21

Query Match 8.7%; Score 224.8; DB 3; Length 3875;  
Best Local Similarity 99.1%; Pred. No. 1.6e-59;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2351 TACGCTGATGCCGATGTTAAGCCAGCCCGCAACCCGCTGACGGCCC 2410  
DB 1482 TGCTTGATGCCGATGTTAAGCCAGCCCGCAACCCGCTGACGGCCC 1541  
QY 2411 TGAAGGAGCTTGTCTGCTCCCGCATCCGCTTACAGCAAGCTGTGACGCTTCCGGAGC 2470  
DB 1542 TGACGGGCTGTCTGCTCCCGCATCCGCTTACAGCAAGCTGTGACGCTTCCGGAGC 1601  
QY 2471 TGATGATGTCAGAGGTTTACCGTCATCACCGGAAACGGCGAGACGAAAGGGCTCGTG 2530  
DB 1602 TGATGATGTCAGAGGTTTACCGTCATCACCGGAAACGGCGAGACGAAAGGGCTCGTG 1661  
QY 2531 ATAGGCTATTTTATAGGTTAATGTCATGATATATATGTTCTTAG 2578  
DB 1662 ATAGGCTATTTTATAGGTTAATGTCATGATATATATGTTCTTAG 1709

Search completed: April 30, 2004, 18:42:53  
Job time: 307.986 secs



Db 181 GATTTCATTGGCGATTCGGGTTGAAAATGGAATAGCGTAATGACTGTGTAATAAAAGTACG 240  
QY 241 ATTCAATGAGTGCAATGCTGCTCTAAATPAAACATGTGCGGATCGGAAAAGCGCAATGA 300  
Db 241 ATTCAATGAGTGCAATGCTGCTCTAAATPAAACATGTGCGGATCGGAAAAGCGCAATGA 300  
QY 301 AGGCACTGATGTAATTTCTTGGGCTACTCTCTAATAAATGCTTAATCTTCTTCACT 360  
Db 301 AGGCACTGATGTAATTTCTTGGGCTACTCTCTAATAAATGCTTAATCTTCTTCACT 360  
QY 361 GCGCAATCAATTCATAGCGCACTCAACATCGGATCAACGATAGAACGGCTGCAAT 420  
Db 361 GCGCAATCAATTCATAGCGCACTCAACATCGGATCAACGATAGAACGGCTGCAAT 420  
QY 421 CATATTGATTTATAGGCAATCAAGATCAATPAACTGCGGATPAAAGGAGAAATAG 480  
Db 421 CATATTGATTTATAGGCAATCAAGATCAATPAACTGCGGATPAAAGGAGAAATAG 480  
QY 481 GCGATGCGCTCATGATGATGCGATCTTGGTTGCGTTGATTCAGTCAATGATAGTAA 540  
Db 481 GCGATGCGCTCATGATGATGCGATCTTGGTTGCGTTGATTCAGTCAATGATAGTAA 540  
QY 541 CAACGTGTGCTAACTTTGGGCGAACATPAAAGTACCTGTGTAAGTTGTCACTTTGT 600  
Db 541 CAACGTGTGCTAACTTTGGGCGAACATPAAAGTACCTGTGTAAGTTGTGTCACTTTGT 600  
QY 601 GACAAACCTAGTCAATGCTTATTTGGCTTATTAATAATGATGATGAGGGGTAAAGA 660  
Db 601 GACAAACCTAGTCAATGCTTATTTGGCTTATTAATAATGATGATGAGGGGTAAAGA 660  
QY 661 CGTAGTCATPAAACAAATTAACAGTACTTGTATCTGATGATGATGTTGTCAAAAGTCT 720  
Db 661 CGTAGTCATPAAACAAATTAACAGTACTTGTATCTGATGATGATGATGTTGTCAAAAGTCT 720  
QY 721 TATTACATTTGACCATCATCATGCACTTAACCTAAATAAGCCGCTGTTTATAGGAA 780  
Db 721 TATTACATTTGACCATCATCATGCACTTAACCTAAATAAGCCGCTGTTTATAGGAA 780  
QY 781 GCGATATGATGTCATCATGATGATGATGTCGTCGCTTGGCGGCAATCTATCCAG 840  
Db 781 GCGATATGATGTCATCATGATGATGATGTCGTCGCTTGGCGGCAATCTATCCAG 840  
QY 841 GTTTTACTGGTAAAGCTCTTAATCCAAATCGGCAAGTTGATGATGATGATGATGATGAT 900  
Db 841 GTTTTACTGGTAAAGCTCTTAATCCAAATCGGCAAGTTGATGATGATGATGATGATGAT 900  
QY 901 GCGGGGATAGTATGACGAAGATGACCGCTCATGCTGAGGAACCTGTGATGAGAT 960  
Db 901 GCGGGGATAGTATGACGAAGATGACCGCTCATGCTGAGGAACCTGTGATGAGAT 960  
QY 961 TTTGATGACGAGACGACGATTTGTGCGCAAAAAGTCATATCTTATCTTAATTTATC 1020  
Db 961 TTTGATGACGAGACGACGATTTGTGCGCAAAAAGTCATATCTTATCTTAATTTATC 1020  
QY 1021 AGGATTCGCTGTGATGAGCAACCCCAACCGGATCGGATGATGATGATGATGATGAT 1080  
Db 1021 AGGATTCGCTGTGATGAGCAACCCCAACCGGATCGGATGATGATGATGATGATGAT 1080  
QY 1081 TCCCATTAAGCTTTATTAACCCGCGTAATGTCAAACAAATAGAGATTTGGTATCAG 1140  
Db 1081 TCCCATTAAGCTTTATTAACCCGCGTAATGTCAAACAAATAGAGATTTGGTATCAG 1140  
QY 1141 CCGGAGGCGCTTAATGCTTGAATCTTCACTTACTCAAGAGAAATGCGCGTGGCT 1200  
Db 1141 CCGGAGGCGCTTAATGCTTGAATCTTCACTTACTCAAGAGAAATGCGCGTGGCT 1200  
QY 1201 TTTGATCATGTGCGCAAAATTCAGACATGCGTGCAGAAATTAAGCGCTGCGTGAATAC 1260  
Db 1201 TTTGATCATGTGCGCAAAATTCAGACATGCGTGCAGAAATTAAGCGCTGCGTGAATAC 1260  
QY 1261 ACATCGGTGATCATTTTCAATGAGAAAGATTTTATGTCGCGATTTATGATGCC 1320  
Db 1261 ACATCGGTGATCATTTTCAATGAGAAAGATTTTATGTCGCGATTTATGATGCC 1320

QY 1321 TAGCCAAATTTGGCTGAGATTAATGCGATGACATTAACCGGCTTGATCAATACC 1380  
Db 1321 TAGCCAAATTTGGCTGAGATTAATGCGATGACATTAACCGGCTTGATCAATACC 1380  
QY 1381 GGGGTGATGCTGATGATCAATTAATGCGGCACTTGTAAAGGCAAGGCAACCA 1440  
Db 1381 GGGGTGATGCTGATGATCAATTAATGCGGCACTTGTAAAGGCAAGGCAACCA 1440  
QY 1441 GCGACGCTTATGCTTGGCAGTCAATGAGTCACTATTTTCAAACTGTTTACGAGGT 1500  
Db 1441 GCGACGCTTATGCTTGGCAGTCAATGAGTCACTATTTTCAAACTGTTTACGAGGT 1500  
QY 1501 TAACTGTTGCAAAATCGTGTACAGTATGATGATGATGATGATGATGATGATGATGAT 1560  
Db 1501 TAACTGTTGCAAAATCGTGTACAGTATGATGATGATGATGATGATGATGATGATGAT 1560  
QY 1561 ATGCAATTAATCTTGTTCATGACCTATATGATGATGATGATGATGATGATGATGATGAT 1620  
Db 1561 ATGCAATTAATCTTGTTCATGACCTATATGATGATGATGATGATGATGATGATGATGAT 1620  
QY 1621 CATGCTGCAATTAATGAGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
Db 1621 CATGCTGCAATTAATGAGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
QY 1681 GATGATGAGGCAATTTTGGCGGCTTGAACGATGCTGATGATGATGATGATGATGATGAT 1740  
Db 1681 GATGATGAGGCAATTTTGGCGGCTTGAACGATGCTGATGATGATGATGATGATGATGAT 1740  
QY 1741 TTAGTGGCAACAGTTTGGGCGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
Db 1741 TTAGTGGCAACAGTTTGGGCGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
QY 1801 TCCGTGAAACCCCTCCATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
Db 1801 TCCGTGAAACCCCTCCATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
QY 1861 GTCAATTAATGAGCAATTAATGAGGTTAGGTTAGGTTAGGTTAGGTTAGGTTAGGTTAGG 1920  
Db 1861 GTCAATTAATGAGCAATTAATGAGGTTAGGTTAGGTTAGGTTAGGTTAGGTTAGGTTAGG 1920  
QY 1921 TTTAAAGAACACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
Db 1921 TTTAAAGAACACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
QY 1981 ATACCTTTAGGCTGCGCAGCTTTTACTGCGGATGATGATGATGATGATGATGATGATGAT 2040  
Db 1981 ATACCTTTAGGCTGCGCAGCTTTTACTGCGGATGATGATGATGATGATGATGATGATGAT 2040  
QY 2041 ACAGTAAAGAAACCAAAATGCTGCGATGATGATGATGATGATGATGATGATGATGATGAT 2100  
Db 2041 ACAGTAAAGAAACCAAAATGCTGCGATGATGATGATGATGATGATGATGATGATGATGAT 2100  
QY 2101 TAGAGAAATGCTTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
Db 2101 TAGAGAAATGCTTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
QY 2161 CCGCTTCGCGGCTTAAAGCCGTTTAACTTCAAGTATGATGATGATGATGATGATGATGAT 2220  
Db 2161 CCGCTTCGCGGCTTAAAGCCGTTTAACTTCAAGTATGATGATGATGATGATGATGATGAT 2220  
QY 2221 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280  
Db 2221 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280  
QY 2281 TCCCATGATTAATTAATGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340  
Db 2281 TCCCATGATTAATTAATGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340  
QY 2341 TCTGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400  
Db 2341 TCTGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400



QY 2401 TGACGCGCCCTGACGCGGCTTGTCTGCTCCGCGCATCCGCTTACAGACAGCTGACCGT 2460  
DB 2401 TACCGCGCCCTGACGCGGCTTGTCTGCTCCGCGCATCCGCTTACAGACAGCTGACCGT 2460  
QY 2461 CTCGCGGAGCTGATGTGTACAGAGGTTTTCACCGTCATCCGAAACGCGGAGACGAA 2520  
DB 2461 CTCGCGGAGCTGATGTGTACAGAGGTTTTCACCGTCATCCGAAACGCGGAGACGAA 2520  
QY 2521 GCGCGCTGATGATGCGCTTATTTTATAGTTATGTCAATGATATATATGTTCTTAA 2578  
DB 2521 GCGCGCTGATGATGCGCTTATTTTATAGTTATGTCAATGATATATATGTTCTTAA 2578

## RESULT 2

US-10-603-260-2  
Sequence 2, Application US/10603260  
Publication No. US20040009570A1  
GENERAL INFORMATION:  
APPLICANT: CJ Corporation  
TITLE OF INVENTION: An alkaline lipase from *Vibrio metchnikovii* RH530 and a  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/603,260  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Koparentin 1.71  
SEQ ID NO 2  
LENGTH: 798  
TYPE: DNA  
ORGANISM: *Vibrio metchnikovii* RH530  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(798)  
OTHER INFORMATION: val11 gene  
US-10-603-260-2

Query Match 31.0%; Score 798; DB 16; Length 798;  
Best Local Similarity 100.0%; Pred. No. 8e-232;  
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 ATGTTTGCACAAAGTCTTATTTACATTTGACATTCATGACATGACCTTAACTAAATAGC 762  
DB 1 ATGTTTGCACAAAGTCTTATTTACATTTGACATTCATGACATGACCTTAACTAAATAGC 60  
QY 763 CCGTTGTTTAAAGGAGAACCATATGATGTCATATGATGATTTGTCGCTCT 822  
DB 61 CCGTTGTTTAAAGGAGAACCATATGATGTCATATGATGATTTGTCGCTCT 120  
QY 823 GCGCGGAATCTATTCAGAGTTTACTGTGAAACGCTCTAATCCAAATCGGCGAATTT 882  
DB 121 GCGCGGAATCTATTCAGAGTTTACTGTGAAACGCTCTAATCCAAATCGGCGAATTT 180  
QY 883 GGTAAATGGGCAATGCTGCGGAGATAGTATGATGACGAATATGACGCGTCAATGTTGA 942  
DB 181 GGTAAATGGGCAATGCTGCGGAGATAGTATGATGACGAATATGACGCGTCAATGTTGA 240  
QY 943 GAACCTGTGATGAGATTTTGTATGACGAGACGACGATTTGTGCGCAAAAAGTCAT 1002  
DB 241 GAACCTGTGATGAGATTTTGTATGACGAGACGACGATTTGTGCGCAAAAAGTCAT 300  
QY 1003 ACTTATCTTATTTTATGAGGATCGCTGCTGATGATGCAACCCCAACCGCATCCGAAT 1062  
DB 301 ACTTATCTTATTTTATGAGGATCGCTGCTGATGATGCAACCCCAACCGCATCCGAAT 360  
QY 1063 GGTGAGGTGATGATTTTCCATTAAGCTTTTAAACCGGTGATGCAACCAATA 1122  
DB 361 GGTGAGGTGATGATTTTCCATTAAGCTTTTAAACCGGTGATGCAACCAATA 420  
QY 1123 GAAGATTTTGTATGACCCGAGCGGCTAATGTTGATCTTACTTACTTAACTTAA 1182  
DB 421 GAAGATTTTGTATGACCCGAGCGGCTAATGTTGATCTTACTTACTTAACTTAA 480  
QY 1183 GAAGAAATGCGCTGCTTTTGTATCATGTGCGCAAAATGACATGCGTGCAGAAAATTA 1242

DB 481 GAAAGAAATGCGCTGCTTTTGTATCATGTGCGCAAAATGACATGCGTGCAGAAAATTA 540  
QY 1243 CCGCTGCGGTTGAATACATCCGCTGATCTATTTTCTTATGAAAAAGGTTTAA 1302  
DB 541 CCGCTGCGGTTGAATACATCCGCTGATCTATTTTCTTATGAAAAAGGTTTAA 600  
QY 1303 GCGGATTTTATGATGCTTACGCGCAATTTGGCGTGAAGTTAATGATGACATTA 1362  
DB 601 GCGGATTTTATGATGCTTACGCGCAATTTGGCGTGAAGTTAATGATGACATTA 660  
QY 1363 CCGCGCTGATGCAATACCGGCGGATGCTGATGACCAATTAATGCGCGCATCTTAA 1422  
DB 661 CCGCGCTGATGCAATACCGGCGGATGCTGATGACCAATTAATGCGCGCATCTTAA 720  
QY 1423 GCGCAAGAGCCCAACACGACCGTTTATGCTCTTCCAGTCAATGATGACATTA 1482  
DB 721 GCGCAAGAGCCCAACACGACCGTTTATGCTCTTCCAGTCAATGATGACATTA 780  
QY 1483 CAACCTGTTTACGAGT 1500  
DB 781 CAACCTGTTTACGAGT 798

## RESULT 3

US-10-603-260-4  
Sequence 4, Application US/10603260  
Publication No. US20040009570A1  
GENERAL INFORMATION:  
APPLICANT: CJ Corporation  
TITLE OF INVENTION: An alkaline lipase from *Vibrio metchnikovii* RH530 and a  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/603,260  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Koparentin 1.71  
SEQ ID NO 4  
LENGTH: 555  
TYPE: DNA  
ORGANISM: *Vibrio metchnikovii* RH530  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(555)  
OTHER INFORMATION: val12 gene  
US-10-603-260-4

Query Match 21.5%; Score 555; DB 16; Length 555;  
Best Local Similarity 100.0%; Pred. No. 7.5e-158;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1561 ATGCAATATTTCTTGTTCATGACCTATATGATGATGCTTGTATGATGATCCGCTTAT 1620  
DB 1 ATGCAATATTTCTTGTTCATGACCTATATGATGATGCTTGTATGATGATCCGCTTAT 60  
QY 1621 CATCGTTCGATTAATTTGGGTTATGTAATCAACCATTAAGTCAACTACTGCTATC 1680  
DB 61 CATCGTTCGATTAATTTGGGTTATGTAATCAACCATTAAGTCAACTACTGCTATC 120  
QY 1681 GATGATGAGGCAATTTTTCGCGCTTACCGATGAGTCACTGACGACCTGCTTATGCT 1740  
DB 121 GATGATGAGGCAATTTTTCGCGCTTACCGATGAGTCACTGACGACCTGCTTATGCT 180  
QY 1741 TTATGTCGACACAGTTTGGCGGATTTGATCAACGTTATCTAAGAAATGCGCGCACCG 1800  
DB 181 TTATGTCGACACAGTTTGGCGGATTTGATCAACGTTATCTAAGAAATGCGCGCACCG 240  
QY 1801 TCGTGTGAACCCCTTCCCATGTGCTCGCATCGGCTCACTTTGCAAGAGCTTCAAT 1860  
DB 241 TCGTGTGAACCCCTTCCCATGTGCTCGCATCGGCTCACTTTGCAAGAGCTTCAAT 300  
QY 1861 GTCAATTAATTAAGCAATTAAGTTTAAAGGTTGACATGATTAATTCAGCAAAATTTGG 1920

Db 301 GTCAATTAATGAGCAATTAGTTAGGGGTGGCACTAGATATTCAGACAAATTTGGG 360  
Qy 1921 TTTAAAGAACACGACGACGATCCGCTATCCAAAAATGAGGAGATATTCAGAGAAAG 1980  
Db 361 TTTAAAGAACACGACGACGATCCGCTATCCAAAAATGAGGAGATATTCAGAGAAAG 420  
Qy 1981 ATACCTTTGGGCTGGGCGACCTTTTACGCGCGATCCAGTCCGATCCGATCCGCTG 2040  
Db 421 ATACCTTTGGGCTGGGCGACCTTTTACGCGCGATCCAGTCCGATCCGATCCGCTG 480  
Qy 2041 ACAGTGAAGAAACCAAAATAGCTGCGATGACGATCATATCCGATATCCACACTTCA 2100  
Db 481 ACAGTGAAGAAACCAAAATAGCTGCGATGACGATCATATCCGATATCCACACTTCA 540  
Qy 2101 TACGAGATGCTGTT 2115  
Db 541 TACGAGATGCTGTT 555

## RESULT 4

US-10-282-122A-41398  
Sequence 41398, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See file wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 41398  
LENGTH: 717  
TYPE: DNA  
ORGANISM: Vibrio cholerae  
US-10-282-122A-41398

Query Match 15.8%; Score 408.2; DB 13; Length 717;  
Best Local Similarity 73.1%; Pred. No. 5.2e-113;  
Matches 524; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

Qy 787 ATGATTTGCTACTATCCATATGATTTGTCGCTCTTGGCCGCAATCTATCCAGTTTAA 846  
Db 1 ATGATTTGCTACTATCCATATGATTTGTCGCTCTTGGCCGCAATCTATCCAGTTTAA 60  
Qy 847 CTGGTGAAGAGCTCTTAATCCAAATCCGCCAGATTTGTATTAATGGGCATTTGCTGGCGG 906  
Db 61 TTGGTCAAGAGCTCTTAATCCAAATCCGCCAGATTTGTATTAATGGGCATTTGCTGGCGG 120  
Qy 907 ATAGTATGACGAAAGATATGACCGCTCATAGTGTGAGAACTGTGCATGAGATTTTGAAT 966  
Db 121 TGGGTGTTATGATGATCTCAGCGCAAGGAGGAGAAACCCGCTGACAAAGATTTTGAAT 180  
Qy 967 GAGCGAGAGAGATTTGTGGCAAAAGTCCATATCTATCTAATTTATCACGAT 1026  
Db 181 TCGCGGCTGGCGCATCTGCCCGCAAAATCATCTTACCCCATCTTCACTAGTAT 240  
Qy 1027 CCGCTGTTGATGGCAACCCCAACCGATCCGAAATGTTGAGTGTAGATTTTCCAT 1086  
Db 241 CTTTGTGAGACGTTATCTTAACCGATCCAAAGCGTTGAGCATAGTATTTCTCAC 300  
Qy 1087 TAGGCTTTAATTAACCGTGAATGTCAACAAATAGAAATTTGTATCCACCCGAG 1146  
Db 301 TAGGCCCTGCTCAATCATTCACACGTACGCAATGAAAGGCTGATGATTAAGAGC 360  
Qy 1147 CGGCTAATGTTGTTGATCTTCACTTACTCAAGAAAGAAATGCGCTGCTGTTGAT 1206  
Db 361 CGTGTATTTGTTGATCTTCACTTACTCAAGAAAGAAATGCGCTGCTGTTGAT 420  
Qy 1207 CATGTGCGCAATTCAGCATGCGTGGCAAAATTAACGCGTGGTGAATACATCC 1266  
Db 421 CATGTGCGCAATTCAGCATGCGTGGCAAAATTAACGCGTGGTGAATACATCC 480  
Qy 1267 GTGGTACTATTTTCAATAGAAAAGATTTTATAGTGGCGATTTATTAATGCTACGCGC 1326  
Db 481 GTAGCTGTCTTCACTTGAAGCTGAGATTTTGTGGCGCATTTATAGCTTACGCGC 540  
Qy 1327 AAATTTGGCGTCGAGTTAATGCAATGACCATTAACGCGCTGATCAATACGCGGCTG 1386  
Db 541 AAATTTGGCGTCGAGTTAATGCAATGACCATTAACGCGCTGATCAATACGCGGCTG 600  
Qy 1387 ATGTCAGTACCAATTAATGCGCATCTTGAAGGAAAGGACCAACGACGACAC 1446  
Db 601 ATGTCAGTACCAATTAATGCGCATCTTGAAGGAAAGGACGACGACGACGACG 660  
Qy 1447 GTTATGCTCTTCCAGTCAAGTCACTAATTTTCAACCTGTTTACGAGTTTA 1503  
Db 661 GTTATGCTCTTCCAGTCAAGTCACTAATTTTCAACCTGTTTACGAGTTTA 717

## RESULT 5

US-10-282-122A-41399/c  
Sequence 41399, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848

Prior Filing Date: 2000-05-23  
Prior Application Number: 60/207,727  
Prior Filing Date: 2000-05-26  
Prior Application Number: 60/230,335  
Prior Filing Date: 2000-09-06  
Prior Application Number: 60/230,347  
Prior Filing Date: 2000-09-09  
Prior Application Number: 60/242,578  
Prior Filing Date: 2000-10-23  
Prior Application Number: 60/253,625  
Prior Filing Date: 2000-11-27  
Prior Application Number: 60/257,931  
Prior Filing Date: 2000-12-22  
Prior Application Number: 60/267,636  
Prior Filing Date: 2001-02-09  
Prior Application Number: 60/269,308  
Prior Filing Date: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
Number of SEQ ID NOS: 78614  
Software: Patent version 3.1  
SEQ ID NO 41399  
Length: 1308  
Type: DNA  
Organism: Vibrio cholerae  
US-10-282-122A-41399

Query Match 9.7%; Score 249.2; DB 13; Length 1308;  
Best Local Similarity 70.6%; Pred. No. 2.4e-64;  
Matches 332; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 23 ACTTCATGCTGTAACCTGCGGCGGCACTTGTGCCAGTGGCGGCGCTAGTCTTCAAG 82  
DB 470 ACCTTCAGTGTAAATCAGAGGCACTCAGGCCAAGCTGACGGCTGTATCTTCAAT 411  
QY 83 ATTAAGGCGCATGATGACCTGTTATTAATGATGATGCTCTCCGACGATCTGAATGCG 142  
DB 410 AGCATGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 351  
QY 143 ATACGAGCTGCGGCTTGTGCGCTTGTGCGGATGCGGATGCGGATGCGGATGCGGAT 202  
DB 350 ATCCGAGTGTGCTGTTTGTGCGCTTGTGCGGATGCGGATGCGGATGCGGATGCGG 291  
QY 203 TGAATAATGGAATAGCGGATGATGATGATGATGATGATGATGATGATGATGATG 262  
DB 290 ACATAATGGAATAGCAAGCATGTCAGAGAGTGTGCTTGAATGATGATGATGATG 231  
QY 263 TCTAATAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 322  
DB 230 GTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 171  
QY 323 CGTACTGCTGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 382  
DB 170 CGATGCGATCAAGAGTGTGCGGATGCTTCACTGCGATGATGATGATGATGATG 111  
QY 383 ACTTCAATCGGATGATCAAGATGATGATGATGATGATGATGATGATGATGATGATG 442  
DB 110 ACCGACATCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 51  
QY 443 AGATCAATTAATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 492  
DB 50 AGATCAATTAATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1

RESULT 6  
US-09-764-847-1872  
Sequence 1872, Application US/09764847  
Patent No. US20020132767A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC009  
CURRENT APPLICATION NUMBER: US/09/764,847  
CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper  
Number of SEQ ID NOS: 2003  
Software: Patent version 2.0  
SEQ ID NO 1872  
Length: 878  
Type: DNA  
Organism: Homo sapiens  
Feature:  
Name/Key: SITE  
Location: (741)  
OTHER INFORMATION: n equals a,t,c, or c  
US-09-764-847-1872

Query Match 8.7%; Score 224.8; DB 9; Length 878;  
Best Local Similarity 99.1%; Pred. No. 5e-57;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGGCTATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2410  
DB 202 TGCTTATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 261  
QY 2411 TGAAGGCTTGTGCTGCTCCGCGATCCGCTTACAGACAGCTGTGACCGTCCGGAGC 2470  
DB 262 TGAAGGCTTGTGCTGCTCCGCGATCCGCTTACAGACAGCTGTGACCGTCCGGAGC 321  
QY 2471 TGATGTGTGAGAGTTTTCACCGTATCACCGAAAGCGCGAGACGAAAGGCGCTGTG 2530  
DB 322 TGATGTGTGAGAGTTTTCACCGTATCACCGAAAGCGCGAGACGAAAGGCGCTGTG 381  
QY 2531 ATACGCTATTTTATAGTAAATGATGATGATGATGATGATGATGATGATGATGATG 2578  
DB 382 ATACGCTATTTTATAGTAAATGATGATGATGATGATGATGATGATGATGATGATG 429

RESULT 7  
US-09-764-891-10029  
Sequence 10029, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
Number of SEQ ID NOS: 10231  
Software: Patent version 2.0  
SEQ ID NO 10029  
Length: 878  
Type: DNA  
Organism: Homo sapiens  
Feature:  
Name/Key: SITE  
Location: (741)  
OTHER INFORMATION: n equals a,t,c, or c  
US-09-764-891-10029

Query Match 8.7%; Score 224.8; DB 10; Length 878;  
Best Local Similarity 99.1%; Pred. No. 5e-57;  
Matches 226; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2351 TAGGCTATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2410  
DB 202 TGCTTATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 261  
QY 2411 TGAAGGCTTGTGCTGCTCCGCGATCCGCTTACAGACAGCTGTGACCGTCCGGAGC 2470  
DB 262 TGAAGGCTTGTGCTGCTCCGCGATCCGCTTACAGACAGCTGTGACCGTCCGGAGC 321  
QY 2471 TGATGTGTGAGAGTTTTCACCGTATCACCGAAAGCGCGAGACGAAAGGCGCTGTG 2530  
DB 322 TGATGTGTGAGAGTTTTCACCGTATCACCGAAAGCGCGAGACGAAAGGCGCTGTG 381

```

Qy      2531 ATAGCCGCTATTTTATAGCTTAATGCATGATAATAATGTTCTTAG 2578
          |||||
Db      382  ATAGCCGCTATTTTATAGCTTAATGCATGATAATAATGTTCTTAG 429

```

## RESULT 8

```

US-09-764-891-10033
; Sequence 10033, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10033
; LENGTH: 878
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (741)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-10033

```

Query Match	8.7%;	Score 224.8;	DB 10;	Length 878;
Best Local Similarity	99.1%;	Pred. No. 5e-57;		
Matches 226; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	235	TAGCGTGAATCCGCAATGTTAAGCCAGAGCCCGACACCCCGCAACACCCGCTGACGGCGCC	241
Db	202	TGCTCTGAATGCCGATGTTAATGAACAGCCCGGACACCCCGCCACACCCGCTGACGGCGCC	261
Qy	2411	TGACGGCGCTGTCTGCTCCCGGCATCCGCTTACAGACAGAGCTGTGACCGTCTCCGGAGAC	2477
Db	262	TGACGGGCTTGTGTGCTCCCGGCATCCGCTTACAGACAGAGCTGTGACCGTCTCCGGAGAC	321
Qy	2471	TGCATGTGTCAAGGTTTTCACCCGTCAACCGAAGCGCGCGAGAGCAAAAGGGCTCTCGTG	2537
Db	322	TGCATGTGTCAAGGTTTTCACCCGTCAACCGAAGCGCGCGAGAGCAAAAGGGCTCTCGTG	381
Qy	2531	ATAAGCGCTATTTTATAGGTTAATGTCATGATTAATAATAGTTCTTAG	2578
Db	382	ATAAGCGCTATTTTATAGGTTAATGTCATGATTAATAATAGTTCTTAG	429

## RESULT 9

US-09-973-278-674  
 ? Sequence 674, Application US/09973278  
 ? Publication No. US20040044191A1  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Fischer et al.  
 ? TITLE OF INVENTION: 123 Human Secreted Protein  
 ? FILE REFERENCE: P2010P2  
 ? CURRENT APPLICATION NUMBER: US/09/973,278  
 ? CURRENT FILING DATE: 2001-10-10  
 ? PRIOR APPLICATION NUMBER: 60/239,899  
 ? PRIOR FILING DATE: 2000-10-13  
 ? PRIOR APPLICATION NUMBER: 09/227,357  
 ? PRIOR FILING DATE: 1999-01-08  
 ? PRIOR APPLICATION NUMBER: ECT/US98/13684  
 ? PRIOR FILING DATE: 1998-07-07  
 ? PRIOR APPLICATION NUMBER: 60/051,926  
 ? PRIOR FILING DATE: 1997-07-08  
 ? PRIOR APPLICATION NUMBER: 60/052,793  
 ? PRIOR FILING DATE: 1997-07-08  
 ? PRIOR APPLICATION NUMBER: 60/051,925  
 ? PRIOR FILING DATE: 1997-07-08  
 ? PRIOR APPLICATION NUMBER: 60/051,929  
 ? PRIOR FILING DATE: 1997-07-08

```

1 PRIOR APPLICATION NUMBER: 60/052, 803
2 PRIOR FILING DATE: 1997-07-08
3 PRIOR APPLICATION NUMBER: 60/052, 732
4 PRIOR FILING DATE: 1997-07-08
5 PRIOR APPLICATION NUMBER: 60/051, 931
6 PRIOR FILING DATE: 1997-07-08
7 PRIOR APPLICATION NUMBER: 60/051, 932
8 PRIOR FILING DATE: 1997-07-08
9 PRIOR APPLICATION NUMBER: 60/051, 916
10 PRIOR FILING DATE: 1997-07-08
11 PRIOR APPLICATION NUMBER: 60/051, 930
12 PRIOR FILING DATE: 1997-07-08
13 PRIOR APPLICATION NUMBER: 60/051, 918
14 PRIOR FILING DATE: 1997-07-08
15 PRIOR APPLICATION NUMBER: 60/051, 920
16 PRIOR FILING DATE: 1997-07-08
17 PRIOR APPLICATION NUMBER: 60/052, 733
18 PRIOR FILING DATE: 1997-07-08
19 PRIOR APPLICATION NUMBER: 60/052, 795
20 PRIOR FILING DATE: 1997-07-08
21 PRIOR APPLICATION NUMBER: 60/051, 919
22 PRIOR FILING DATE: 1997-07-08
23 PRIOR APPLICATION NUMBER: 60/051, 928
24 PRIOR FILING DATE: 1997-07-08
25 PRIOR APPLICATION NUMBER: 60/055, 722
26 PRIOR FILING DATE: 1997-08-18
27 PRIOR APPLICATION NUMBER: 60/055, 723
28 PRIOR FILING DATE: 1997-08-18
29 PRIOR APPLICATION NUMBER: 60/055, 948
30 PRIOR FILING DATE: 1997-08-18
31 PRIOR APPLICATION NUMBER: 60/055, 949
32 PRIOR FILING DATE: 1997-08-18
33 PRIOR APPLICATION NUMBER: 60/055, 953
34 PRIOR FILING DATE: 1997-08-18
35 PRIOR APPLICATION NUMBER: 60/055, 950
36 PRIOR FILING DATE: 1997-08-18
37 PRIOR APPLICATION NUMBER: 60/055, 947
38 PRIOR FILING DATE: 1997-08-18
39 PRIOR APPLICATION NUMBER: 60/055, 964
40 PRIOR FILING DATE: 1997-08-18
41 PRIOR APPLICATION NUMBER: 60/056, 360
42 PRIOR FILING DATE: 1997-08-18
43 PRIOR APPLICATION NUMBER: 60/055, 684
44 PRIOR FILING DATE: 1997-08-18
45 PRIOR APPLICATION NUMBER: 60/055, 984
46 PRIOR FILING DATE: 1997-08-18
47 PRIOR APPLICATION NUMBER: 60/055, 954
48 PRIOR FILING DATE: 1997-08-18
49 PRIOR APPLICATION NUMBER: 60/058, 785
50 PRIOR FILING DATE: 1997-09-12
51 PRIOR APPLICATION NUMBER: 60/058, 664
52 PRIOR FILING DATE: 1997-09-12
53 PRIOR APPLICATION NUMBER: 60/058, 660
54 PRIOR FILING DATE: 1997-09-12
55 PRIOR APPLICATION NUMBER: 60/058, 661
56 PRIOR FILING DATE: 1997-09-12
57 NUMBER OF SEQ ID NOS: 947
58 SOFTWARE: Patentin Ver. 2.0
59 SEQ ID NO 674
60 LENGTH: 878
61 TYPE: DNA
62 ORGANISM: Homo sapiens
63 NAME/KEY:
64 FEATURE:
65 LOCATION: (741)..(741)
66 OTHER INFORMATION: n equals a,t,g, or c
67 US-09-973-878-674

```

Query Match	8.7%	Score 224.8;	DB 13;	Length 878;
Best Local Similarity	99.1%;	Pred. No. 5e-57;		
Matches 226;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
2351	TACGCTGATGCCGATAGTTAAGCCAGCCGACACCCGCCACACCCGCTGACGCGCCC	2410		

Db	202	TGCTGTATGCGCATAGTTAAGCCAGCCCCCAACCGGCCAAACCCGCTGACCGGCC	261
Qy	2411	TGACGGGGATTGTCGTCTCCCGGCATCCGGTTACAACAAGCTGTAAACGCTCTCCCGGGAGC	2476
Db	252	TGACGGGCGTTGTCTCTCCCGGCATCCGTTACACACAGCTGTAAACGCTCTCCCGGGAGC	321
Qy	2471	TGCATGTGTCAAGGTTTTCAACCGTCATCACCGAAACGGCGCAGACGAAGGGCCTCGTG	2533
Db	332	TGCATGTGTCAAGGTTTTCAACCGTCATCACCGAAACGGCGCAGACGAAGGGCCTCGTG	381
Qy	2531	ATAGCCATTTTTTATAGGTTAATGTCATGATTAATATAGTTCTTAG	2578
Db	382	ATAGCCATTTTTTATAGGTTAATGTCATGATTAATATAGTTCTTAG	429

RESULT 10  
US-10-092-154-1872  
; Sequence 1872, Application US/10092154  
; Publication No. US20030054375A1  
JOURNAL INFORMATION

1 APPLICANT: Rosen et al.  
 2 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 3 FILE REFERENCE: PC00901  
 4 CURRENT APPLICATION NUMBER: US/10/092,154  
 5 CURRENT FILING DATE: 2003-03-07  
 6 NUMBER OF SEQ ID NOS: 2003  
 7 Prior Application removed - See File Wrapper or Palm  
 8 SOFTWARE: PatentIn Ver. 2.0

```

:      TYPE: DNA
:      ORGANISM: Homo sapiens
:      FEATURE:
:      NAME/KEY: misc_feature
:      LOCATION: (741)
:      OTHER INFORMATION: n equals a,t,g, or c
:      OS-10-032-154-1872

```

[illegible]

```

RESULT 11
US-09-758-987-6/c
Sequence 6, Application US/09758987
Patent No. US20010029014A1
GENERAL INFORMATION:
APPLICANT: De Beuckeleer, Marc
TITLE OF INVENTION: Methods and kits for identifying elite event GAT-ZM1 in biological
FILE REFERENCE: 514412-2025.1
CURRENT APPLICATION NUMBER: US/09/758,987
CURRENT FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: 09/481,049
PRIOR FILING DATE: 2000-01-11

```

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1073)
; OTHER INFORMATION: Sequence comprising a 5' flanking region of GAT-ZM1
US-09-758-987-6

```

Query Match	8.7%;	Score 224.8;	DB 9;	Length 1073;
Best Local Similarity	99.1%;	Pred. No. 5.8e-57;		
Matches 226;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

QY 2411 TGACGGGCTTGTCTGCTCCCGGCATCCGCTTTACAGACAAAGCTGACCGCTCCCGGAGC 2470

Db 649 TGAAGGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAAGCTGACCGCTCCCGGAGC 590

QY 2471 TGCATGTGTCAAGAGTTTACCGGTCACTACCGAAAGCGCGAGACCAAAAGGCTCGTG 2530

Db 589 TGCATGTGTCAAGAGTTTACCGGTCACTACCGAAAGCGCGAGACCAAAAGGCTCGTG 530

QY 2531 ATAGGCCATTTTATAGTTAATAGTATAGTATATTAATAGTTCTTAG 2578

Db 529 ATAGGCCATTTTATAGTTAATAGTATAGTATATTAATAGTTCTTAG 482

```

RESULT 12
US-09-891-865A-11
; Sequence 11, Application US/09891865A
; Publication NO. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26S
; CURRENT APPLICATION NUMBER: US/09/891,865A
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: M98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2297
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of cloning vector
; OTHER INFORMATION: derived from pUC18
; US-09-891-865A-11

```

Query Match	8.7%	Score 224.8	DB 10	Length 2297
Best Local Similarity	99.1%	Pred. No. 1e-56		
Matches 226	Conservative 0	Mismatches 2	Indels 0	Gaps 0
Qy	2351	TACGCTATGCCGAGATGTTAAGCAGACCCCGCAGACACCCGCTGACGGGCC	2410	
Db	133	TGCTCTGATGCCGATAGTTAAGCAGACCCCGCAGACACCCGCTGACGGGCC	192	
Qy	2411	TGACGGGCTTGTCGCTCCCGGCATCCGCTTACAGACAAGTTGACCGTCTCCGGAGC	2470	
Db	193	TGACGGGCTTGTCGCTCCCGGCATCCGCTTACAGACAAGTTGACCGTCTCCGGAGC	252	
Qy	2471	TGCATGTGTACAGAGTTTACACGCTATCACCCGAGACGGCGGAGACGAAAGGCGCTGTG	2530	
Db	253	TGCATGTGTACAGAGTTTACACGCTATCACCCGAGACGGCGGAGACGAAAGGCGCTGTG	312	

QY 2531 ATACGCCATTTTATAGTTAATGCATGATAATAATGTTCTTAG 2578  
DB 313 ATACGCCATTTTATAGTTAATGCATGATAATAATGTTCTTAG 360

### RESULT 13

US-10-365-062-5/c  
Sequence 5, Application US/10365062  
Publication No. US20030145343A1  
GENERAL INFORMATION:  
APPLICANT: Pfizer Inc.  
APPLICANT: McNeil, John D.  
TITLE OF INVENTION: TRANSGENIC ANIMALS EX-  
FILE REFERENCE: PC10142B  
CURRENT APPLICATION NUMBER: US/10/365, 062-5  
CURRENT FILING DATE: 2003-02-11  
PRIOR APPLICATION NUMBER: US 09/496,445  
PRIOR FILING DATE: 2000-02-02  
PRIOR APPLICATION NUMBER: US 60/118,478  
PRIOR FILING DATE: 1999-02-03  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 5  
LENGTH: 2462  
TYPE: DNA  
ORGANISM: Escherichia coli  
US-10-365-062-5

Query Match	8.7%	Score 224.8	DB 15	Length 2462
Best Local Similarity	99.1%	Pred. No. 1.1e-56		
Matches 226, Conservative	0	Mismatches 2	Indels 0	Gaps 0

Qy	2411	TGACGGGCTTGTCTCTCCCGGATCGGCTTACAGACAAGCTGTGACCGTCTCCCGGAGC	2470
Db	2298	TGACGGGCTTGTCTCTCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCCGGAGC	2239
Qy	2471	TGCATGTGTCAAGGTTTTCACACCGTCATCACCGAACAACCGCGAACAAGGCGCTCGTG	2538
Db	2238	TGCATGTGTCAAGGTTTTCACACCGTCATCACCGAACAACCGCGAACAAGGCGCTCGTG	2179
Qy	2531	ATAGCCGTAATTTTATAGGTAAATGCATGATTAATATAGTTTCTTAG	2578
Db	2178	ATACCCCTAATTTTATAGGTAAATGCATGATTAATATAGTTTCTTAG	2131

RESULT 14

```

US-10-244-142A-2/C
Sequence 2, Application US/10244142A
Publication No. US20030199516A1
GENERAL INFORMATION:
APPLICANT: Moser, Heinz B.
APPLICANT: Baird, Eldon E.
APPLICANT: Buerli, Roland W.
APPLICANT: Ge, Yigong
APPLICANT: White, Sarah
APPLICANT: Genesoft, Inc.
TITLE OF INVENTION: Methods of Treating Infection by Drug Resistant
TITLE OF INVENTION: Bacteria
FILE REFERENCE: 020891-00910US
CURRENT APPLICATION NUMBER: US/10/244,142A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/222,704
PRIOR FILING DATE: 2001-09-13
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2762
TYPE: DNA

```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:double stranded
; OTHER INFORMATION: circular Plasmid B
US-10-244-142A-2

```

Query Match	8.7%;	Score 224.8;	DB 15;	Length 2762;
Best Local Similarity	99.1%;	Pred. No. 1.2e-56;		
Matches 226;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

QY 2351 TACCTCTATGCCGATAGTTAAAGCCACCCCGACACCCCGCAACACCCGCTGACGCGCC 2410

Db 2410 TGCTCTATGCCGATAGTTAAAGCCACCCCGACACCCCGCAACACCCGCTGACGCGCC 2351

QY 2411 TGACGGGCTGTCTGCTCCCGGCATCCGCTTACAGACATAGTGTGACCGTCTCCGGAGC 2470

Db 2350 TGACGGGCTGTCTGCTCCCGGCATCCGCTTACAGACATAGTGTGACCGTCTCCGGAGC 2291

QY 2471 TGCATGTGTCAAGAGTTTTCACCGTCATTCACCCGAAGCGCGAGACGAAGAAGGCGCTGTG 2530

Db 2290 TGCATGTGTCAAGAGTTTTCACCGTCATTCACCCGAAGCGCGAGACGAAGAAGGCGCTGTG 2231

QY 2531 ATACGCCATATTTATAGTTTAAGTCATGATATATAATAGCTTCTTGA 2578

Db 2230 ATACGCCATATTTATAGTTTAAGTCATGATATATAATAGCTTCTTGA 2183

**RESULT 15**

```

US-10-244-142A-1/C
/ Sequence 1, Application US/10244142A
/ Publication No. US20030199516A1
/ GENERAL INFORMATION:
/ APPLICANT: Moser, Heinz E.
/ APPLICANT: Baird, Eldon B.
/ APPLICANT: Buerli, Roland W.
/ APPLICANT: Ge, Yigong
/ APPLICANT: White, Sarah
/ APPLICANT: Genesoft, Inc.
/ TITLE OF INVENTION: Methods of Treating Infection by Drug Resistant
/ TITLE OF INVENTION: Bacteria
/ FILE REFERENCE: 020891-00910US
/ CURRENT APPLICATION NUMBER: US/10/244,142A
/ CURRENT FILING DATE: 2002-09-12
/ PRIOR APPLICATION NUMBER: US 60/322,704
/ PRIOR FILING DATE: 2001-09-13
/ NUMBER OF SEQ ID NOS: 20
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 2803
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:double stranded
/ OTHER INFORMATION: circular Plasmid A
US-10-244-142A-1

```

Query Match	8.7%;	Score 224.8;	DB 15;	Length 2803;
Best Local Similarity	99.1%;	Pred. No. 1.2e-56;		
Matches 226;	Conservative	0;	Mismatches 2;	Indels 0;
				Gaps 0;

[illegible]

Mon May 3 09:19:46 2004

us-10-603-260-1.rnpb

Page 9

Db 2240 ATACGCTATTATTATAGGTAAATGTCATGATPATATAGTTCTTAG 2193

Search completed: April 30, 2004, 23:19:29  
Job time : 1149.07 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 08:18:22 ; Search time 3218.8 Seconds

(without alignments)  
10745.552 Million cell updates/sec

Title: US-10-603-260-2

Perfect score: 798  
Sequence: 1 atgttgcacaaagctcta.....ttcaacctgttaccagagc 798

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mn:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_hcg\_hum:\*  
39: em\_hcg\_mus:\*  
40: em\_hcg\_other:\*  
41: em\_hcg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	434.2	54.4	10977	1	AE004352
C 2	340.6	42.7	241900	1	AP005337
C 3	339	42.5	306147	1	AE016805
C 4	282.8	35.4	303450	1	AP005085
C 5	42	5.3	1141	6	AX083744
C 6	40.2	5.0	146953	2	AC138591
C 7	40	5.0	2067	3	AF146758
C 8	39.2	4.9	16237	2	AC012853
C 9	39.2	4.9	74085	3	AC012650
C 10	39.2	4.9	108350	3	AC011761
C 11	39.2	4.9	302540	3	AE003571
C 12	39	4.9	128583	2	AC135793
C 13	39	4.9	168422	2	AC119922
C 14	39	4.9	187063	2	AC107808
C 15	38.8	4.9	2669	2	AC125495
C 16	38.8	4.9	304680	1	AE017002
C 17	38.4	4.8	303250	1	AP003193
C 18	38.2	4.8	1318	9	BC022446
C 19	38.2	4.8	5533	9	HS0807994
C 20	38.2	4.8	5841	9	HS0803639
C 21	38.2	4.8	6242	9	HS0807593
C 22	38.2	4.8	110146	2	AC022421
C 23	38.2	4.8	160142	2	AC093684
C 24	38	4.8	2000	6	AX655393
C 25	37.6	4.7	2000	6	AX655393
C 26	37.4	4.7	1141	6	AX083744
C 27	37.2	4.7	189271	2	AC129309
C 28	37	4.6	17731	2	AC129133
C 29	37	4.6	187121	2	AC112418
C 30	36.8	4.6	115762	2	AP000663
C 31	36.8	4.6	155332	2	AC109581
C 32	36.8	4.6	172830	2	AP001330
C 33	36.8	4.6	186955	9	AP000848
C 34	36.6	4.6	4894	9	HS0807990
C 35	36.6	4.6	221640	2	BX322637
C 36	36.4	4.6	82073	9	HS172K10
C 37	36.4	4.6	168253	2	AC068527
C 38	36	4.5	73455	9	HS256M13
C 39	36	4.5	161514	9	AC078845
C 40	36	4.5	166239	9	HS214G14
C 41	36	4.5	255339	2	AC112075
C 42	36	4.5	273908	2	AL162851
C 43	36	4.5	340000	9	HS21C004
C 44	35.8	4.5	254050	3	PFA929358
C 45	35.6	4.5	552	6	AR345675

#### ALIGNMENTS

RESULT 1	AE004352/c	10977 bp	DNA	linear	BCT 10-OCT-2003
LOCUS	AE004352				
DEFINITION	Vibrio cholerae O1 biovar eltor str. N16961				
ACCESSION	AE004352				
VERSION	AE004352.1				
KEYWORDS	GI:9657475				
SOURCE	Vibrio cholerae O1 biovar eltor str. N16961				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.				
REFERENCE	1 (bases 1 to 10977)				
AUTHORS	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, R.K., Peterson, J.D., Umayam, L.,				

**TITLE**  
 DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*  
**JOURNAL**  
 Nature 406 (6795), 477-483 (2000)  
**PUBMED**  
 20406833  
**REFERENCE**  
 10952301  
**AUTHORS**  
 2 (bases 1 to 10977)  
 Heideisberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,  
 Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,  
 Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,  
 Esmolaeva,M.D., Vamathevan,J., Bess,S., Qin,H., DiGioli,I.,  
 Sellers,P., McDonald,L., Utechtack,T., Fleischmann,R.D.,  
 Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,  
 Mekalanos,J.J., Venter,J.C. and Fraser,C.M.  
**TITLE**  
 Direct Submission  
**JOURNAL**  
 Submitted (14-JUN-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
**FEATURES**  
 Location/Qualifiers  
 source  
 1..10977

gene  
complement (81. .974)  
/gene="VCA0090"  
complement (81. .974)  
/gene="VCA0090"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAFP6004.1"  
/db\_xref="GI:9657476"  
/translation="MSLVNLIQLCLAIWNGSFLFWRIMASFGPAVILEARGPAA  
LSFLVPAQLRRSLPIRQWRRFFILGILINRVAPFLFAYALILNYSLSINSTPR  
LWGAIVIGLWMTGFLSRRAVAGLLIGVGVAIVGMDVVALIGHAALPMVCAVALAAS  
YGLANVYIGKMQPOLSAFENAGSNMAALCWAFLPMFVPLAEPTSSLEMGVALIGVIT  
CTGLAVLYLFLRVKAIAGAASLTVFLIPVFGILMGVWIDPEIGLNTLFTCLIVLAG  
TLVLTGFSRPNRLSRQPPH"  
1204. 1500  
/gene="VCA0091"  
1204. 1500  
/gene="VCA0091"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAFP6005.1"  
/db\_xref="GI:9657477"  
/translation="MARPKPIRRICHPSPACFPKRPVPIQLARVELAPDEHEAALL  
VDQLQDQQAALQVSRQTLANLVKARKFVVDCLLHQALYIAIDNRKSD"  
complement (1671. .2300)  
/gene="VCA0092"  
complement (1671. .2300)  
/gene="VCA0092"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAFP6006.1"  
/db\_xref="GI:9657478"  
/translation="MKIVILHGLWNGHGLVMPLOAQRANKLGYQTEVISYNTLADDERK  
VFOITDLAQRINVLVNGSHGIMIGYRSRSPNYSHYVALASPKGASIV  
KIOGLIGLAKGNALVGLQILOSWSWELPQSLGICAGTLRGFRPIILGSGMDGVIV  
TVAEQIISGMDPHLLHSHSGIVYSHNTAQQIDYFIHNPQFKKIPK"  
2400. .2573  
/gene="VCA0093"

gene 2400. 2573  
/gene="VCA0093"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAP96007.1"  
/db\_xref="GI:9657479"  
/translation="MEQPCMLRSVPIITWKRYVNEDERTCLIPTPSRLLAKGLEKHTH  
DDCLFKSKSDLM"  
2801. 3157  
/gene="VCA0094"  
2801. 3157  
/gene="VCA0094"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAP96008.1"  
/db\_xref="GI:9657480"  
/translation="MDVNIKSPRIPIINDVALQOLERISTKMTWTHLMHLDSSSARF  
GELGYSLNGISIAKMLSHRLKLEBIGVMWREQGTIVVLYYGLTSGRVANGLYLM  
IDRVKNEREPLELSN"  
3426. 4124  
/gene="VCA0095"  
3426. 4124  
/gene="VCA0095"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAP96009.1"  
/db\_xref="GI:9657481"  
/translation="MYQKLSRMFLFSVVLFEGISILATYTSSTPTNLIPEROGH  
VIARQPEPTEISHEHPAYLEISTEKRYVNEOGEVQCEAPRVIEPOKLEQVLEPI  
NPTPEIDILKRSQALVMDIOAIPEPNIIDORRYLMAGSALPLPNESIRIVLDEQOT  
QAQGIQRTAGNFERVQGGTFPTFTFGSHSINGFIGHQSSIKIEGVHQWVYQVPS  
HGFEHDVQSHDRH"  
4167. 5390  
/gene="VCA0096"  
4167. 5390  
/gene="VCA0096"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAP96010.1"  
/db\_xref="GI:9657482"  
/translation="WKKASLALAVSGAGSPATWATTEIDVGLGYPTDAKGFQKEHV  
ACOMHSVNVANKYKDCGLDITKYLATATKCYOVQDQGLQSEVDAITPNNRIDP  
AFADVEARQVQCADMWATIRYIDVNNSPYERQPNQSYISGLAMATVAPAMKYPD  
NAKSMVSHSLTNECGAETFIHEIGHNFGAHAEQYRELFPHNGTBEVDAYGIGIG  
QFATIMVPHLFGVGRSVKSSPNLQCEGAPCGVKDANSVRAIIGLTAPIHAYVYTS  
KPPVDDGPGGTEPTDNTNNVFTIKGLALPDKMLTLPLVPSAQOSTQAVIDITH  
EYRDLSTRLAPDGSYVVKQANRYDRGOSYVQVFLTNDVDPSSAAEGEWRLIEODHF  
GGKLTGNOFOITFP"  
5403. 6119  
/gene="VCA0097"  
5403. 6119  
/gene="VCA0097"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAP96011.1"  
/db\_xref="GI:9657483"  
/translation="MIVTIDMICRLAPKGIETLLVRSNPNRPGCGMALIPGWWFD  
EDLAAQGEEDPDQFDSARRICRQKITHYENFISDVLQGNRPDQSGSISISHA  
VLNHSNVTQLEBAEMSKRVNWFALERTILQGVQLAFDHAQIQLAMKCLRAVEYYS  
VLSRLKEFIVADIIEAYKAFGVDMVMTKSRLLIDTGIVISTNKILASKQGGKGP  
ATVYRLASNEITYQGTCLRG"  
6518. 7825  
/gene="VCA0098"  
/gene="VCA0098"

	6518 . 7825	/gene="VCA0099"
	/note="similar to GB:J05568 SP:P18133 PID:147307 GB:U00096	
	PID:I651456; identified by sequence similarity; putative"	
	/codon start=1	
	/transl_table=1	
	/product="nicotinate phosphoribosyltransferase"	
	/protein_id="AAP96012.1"	
	/db_xref="GI:96574848"	
	/translation="MNPRLPEPHIIRSLDLDAKIMMQAIHFPYDVSRYELLYR	
	SEESASGLDRIROEIAHLGSTRSDIDHTLIOAHFLKATPLQSLRHFFPOEV	
	EMGVXGGGQQLRISIRGSWRDTILYETLMALVSVSRQRVAEPADLPVKLT	
	KLDLKPIERRGGINPSLTEMGRRRPSSOVCDVACIKOEIPVCULTSNHFAR	
	BFDLQTEIGTIAHEWFMGOALVNBRDSQQVALLEMLIAFDLITIDAFAFN	
	DPNHNLANYDVNRDSCGPWMGDGMKLHAHQCLGIDPTTGFLFESGIDPDOLLEC	
	EYPNGRVKISRIGCTPEFLMDLNARKNAAGVEYRELSTVIKAECSGPAKISDPKK	
	AMCEDPIFLANLRKRFINIEDVDALLIQELHRQRSPPHYISAA"	
gene	7856 . 8896	/gene="VCA0099"
CDS	7856 . 8896	/gene="VCA0099"
	/note="similar to GB:U00096 PID:1742678 PID:1742666	
	PID:1742694 PID:1742718; identified by sequence	
	similarity; putative"	
	/codon_start=1	
	/transl_table=1	
	/product="oxido-reductase, gfo(1dh/MocA family)"	
	/protein_id="AAF96013.1"	
	/db_xref="GI:9657485"	
	/translation="MNPRLPKVGIVGYYSKATHLPIDITESTLSATISSSQNVV	
	QEDPQIAPDSAOPLTESDVDLVITAFNNVHPPLAKALEHGKAVITYEKRPVTQI	
	EOGTLLIELAQOGLISVFENRRWDQDFVLKKLIQSGLDGVKVESHSFRPVV	
	RORREDOAEDEGGGILLFDLAPLDQAOLFPLSADCCMKRPDATTIDYDPLQDY	
	YPPVRVRAHLNIAPSPENVRYQVIGSLTKYKGLDPEQDELKGESEPTHOWSOENP	
	BOYLGHAEENENVITFLGSGYHFTQVVVAIRNGASNPFVASBALQSILTALAE	
	SSAAGQRLAVTL"	
gene	complement (8933 . 9304)	
CDS	/gene="VCA0100"	
	complement (8933 . 9304)	
	/gene="VCA0100"	
	/note="similar to GP:7226016; identified by sequence	
Query Match	54.4%; Score 434.2; DB 1; Length 10977;	
Best Local Similarity	72.5%; Pred. No. 8.2e-117;	
Matches	576; Conservative 0; Mismatches 218; Indels 1; Gaps 1;	
DB	3 GTTGTCAACAAGCTTTATTACATTTGACCATCATCATGCGACTTACCTTAATAAGGCC 62	
6200	GTTATTCACATTTTGTATTACATTTGACCTTAGCATTAGTACTTACCTGAAT-AGCG 6142	
OY	63 GTTGTTATTAGGAAGCCATTGATTTGTCACTATGCATATGATTTGTCTGCCTTTGC 122	
Db	6141 CACCTAATGTAGAACCTGTATGATTTGTCACTATGATATGATCTGTTGGCGCTTGC 6082	
OY	123 GCCGAATCTATCCGAGTTTACTGGTAAAGCGCTTAATCCAATCGGCAGATTGTG 182	
Db	6081 CCCGAGGGCATTGAGACTCTGTGTCTCAAAGCAGCAATCCAAAGCCCTGATTGGG 6022	
OY	183 TAAATGGCATTTGCTGGCGGAGATAGTATGACGAGAATATGACCGCTCATGTGTGAGA 242	
Db	6021 TCTTTGGCGCATCCCCGGTGTGGGTGTTGATATGAGATCTCAGCGCACAAAGGTGGGA 5962	
OY	243 ACGCTGATGAGATTTTGATGACGAGACGAGATTTTGTGGGCAAAAAGTCCATAC 302	
Db	5961 ACCGCTGACAAAGTTTTGATTCGCGCGGTGGCGCATCTGCGCCGAGAAATCCATAC 5902	
OY	303 TTATCTAATTTTATCAGCGATCGCTGTTGATGGCAACCCCAAGCGATCCGATGG 362	
Db	5901 TTACCCCACCTTCATCAGTGATCCCTTGTGGACGATATCTTAAACCGCATCCAGCG 5842	
OY	363 TTGAGTGTGATTTTCCCACTTAGCGCTTATTAACCCCGGATATGCAACAATAAGA 422	
Db	5841 TTGGAGCATGATTTTCTACTATGCGCCTGTCTATATTCATTCACGTCACGCGATTTGA 5782	

Oy		423	AGATTTTGGTATGCAACCCCGAGCGCCTTAATGGTTTGATCTTCATCTTTACTCAAAAGA	482
Db		5781	AGAGCGTGAATGATAAAGAGCGGTGTAAATTTGGTTTGCATGAGACGATTTTGCAAG	5722
Oy		483	AGAAATGCCCGTGGCTTTTGTATCATGTGCGGCAATTCAGCATGCGTGGCAAAATTAACG	542
Db		5721	TAAACAAGTTCTGGCCGTTTATCATGTGGGCAAAATTCAGCATGATGAAAAAACGCGC	5662
Oy		543	CGCTGCGGTTGAATATCACATCCGNGTACTATTTTCTTAGAAAAAGATTTTATGTGCG	602
Db		5661	CGCAGCGGTTGAGTACACCTCTGTAGTGCCTTCTCATCTTGACGTGAAATTTTTGTGGCG	5602
Oy		603	GGATATTATTGATGTCCTTACGCCCAAATTTGGCGCTGGAATTAAATGCAATGACATTAACG	662
Db		5601	GGAATATTATTGAGGCTTACCGCAAAATTTGGAAGTGAAGTCAATGCGATGACGGTAAACG	5542
Oy		663	CCGCTTATATCAATATCCGGGGGNGATGCGTCAATCAATTAATAATGGCGCATCTTGTAAGG	722
Db		5541	TGCGCTATATGATACC GG GTTAATGTCAATGACAAATTAATTTGCCGCTCTAGCAAAAG	5482
Oy		723	CAAAAGAGCCAAACCAACCCAGTGTATGCTTCCAGTCATGAAATGCACTTAATTTTCA	782
Db		5481	TAAAGGCGGCAAAACCTGCGACGGTTTATCCCTTAGACAGTAACGAATGCAATCTAATTTTCA	5422
Oy		783	AACCTGTTTACGAGG	797
Db		5421	GACTTGCTTGCGTGG	5407
RESULT 2				
AP005337/c		241900 bp	DNA	linear BCT 04-DEC-2003
LOCUS				
DEFINITION		vibrio vulnificus XJ016 DNA, chromosome I, complete genome, section		
ACCESSION		AP005337 BA000037		
VERSION		AP005337.1 GI:37198635		
KEYWORDS				
SOURCE		Vibrio vulnificus XJ016		
ORGANISM		Vibrio vulnificus XJ016 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.		
REFERENCE		1		
AUTHORS		Chen,C.Y., Wu,K.M., Chang,Y.C., Chang,C.H., Tsai,H.C., Liao,T.L., Liu,Y.M., Chen,H.J., Shen,A.B., Li,J.C., Su,T.L., Shao,C.P., Lee,C.T., Hor,L.I. and Tsai,S.F. Comparative Genome Analysis of Vibrio vulnificus, a Marine Pathogen Genome Res. 13, 2577-2587 (2003) 2 (bases 1 to 241900)		
JOURNAL		Chen,C.Y., Wu,K.M. and Tsai,S.F. Direct Submission Submitted (29-May-2002) Shih Feng Tsai, National Health Research Institutes, Division of Molecular and Genomic Medicine; 129, Yen-Chiu-Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China (E-mail:petsai@nhri.org.tw, Tel:886-2-8146-1041, Fax:886-2-2789-0484)		
COMMENT		This sequence was determined by the Sequencing Core of the National Yang-Ming University Genome Research Center (YMGC; http://genome.ym.edu.tw).		
FEATURES		Location/Qualifiers		
source		1..241900		
gene		/organism="Vibrio vulnificus XJ016"		
CDS		/mol_type="genomic DNA" /strain="XJ016" /db_xref="taxon:196600" /chromosome="I" 197..1024 /gene="VV1706" 197..1024 /gene="VV1706" /codon_start=1 /transl_table=11 /product="formate dehydrogenase accessory protein"		

/protein\_id="BAC94470.1"  
/db\_xref="GI:37198636"  
/translation="MKPKVITKSNPLQTEVEVPEDEGKLIKQICAPRLTYLAK  
KWEVTLMTLGSREPALVUGYIKKQSPFSEVDSLSVITDMETNSAAVYTKENVHIE  
QALKKVTVSGCGQGTVMGNVKNQLEHYTPQVTLKQSTLYTLEALHYNDYTKVAG  
AYGCAKCVGKDEVRSEFVEDGRNNAVDTLAGMNLNDSGADKIPTYTGRLTSVIVIK  
VVMQGPVLLSSSGVTVQMGLDLARQGITTIARAGLRFQVPTGGENIIFVKGQES"  
1151..2908  
/gene="VV1707"  
1151..2908  
/gene="VV1707"  
/transl\_table=1  
/product="methyl-accepting chemotaxis protein"  
/protein\_id="BAC94471.1"  
/db\_xref="GI:37198637"  
/translation="MSGSPVQKRFPSFISRAKLLINVTLLGVAAVAYENNSAK  
LSTLSEHSTENSSSIDLMLRHEKDYIARSEKLTSEFATYSQDQIDHLSIDL  
ASSSIAIDQROQIVTTLQYOQFOHQAOLANDALQSEDLFAARSMKTNVLSAND  
GELTAQFLDLHDLTFTVEPTQEHKVALBGLSQFAKEADLPAPFDYSVLSRVY  
SAIETGLTANESLFGALRSNHYKTEQATTDQTTIKQGEQTADIRSHQOMGVYI  
VVLSSLYVIGRSILSRIRKAINLMDIANGSGDLYVMNKGSDLAOLSRDLF  
TISHNINIKOLASVMDVLCSSGSSSQAORSMONARVNSPADALQLTATGRID  
TLASIEESQAOIQALEAOSRENNAYSAIQIAECTNLLANALIEAPAGESERG  
AAVADVRQLSRITNDSTQISTVQALVYKIAQVKNKADSDPARITNDTRVIE  
ALEIGSTQITEMDNTQIATASEQSWASBIDRNIQTQIADQDYEIVSGSVCS  
EQVDVSHLEKTVAAQFKY"  
complement (3018..4610)  
/gene="VV1708"  
complement (3018..4610)  
/gene="VV1708"  
/note="containing duplicated ATPase domains"  
/codon\_start=1  
/transl\_table=1  
/product="ABC-type transport system, ATPase component"  
/protein\_id="BAC94472.1"  
/db\_xref="GI:37198638"  
/translation="WISTANTTQFGAKPLPENISYKSGSNRYGLIGANGCKSTM  
KILSGLEPTSGVSYDPNERVAKLNQDPAVEEFTVIDVTIMGHELMKVOEEDRI  
YSLPENSEBQKRVADLEVEFAMDMGYMAKAGELLAVGI.PMEOHGLMNSVAPG  
KRLVLAQILFADPDIMLDEPTNMDMTIRLLETTIARNCWTMIISHDHPLANSV  
CTAMADLDGELVPGNDEVTWATASOARELISPNACKKQIAELCTGFVRFANA  
SKKQATSAKQIDKIKLDEVASRSONPFRFESKLEFRNALIVENLSOEFEDL  
SDPNALPEYGERVAILTGENGVKTTILNTLAVLERPTEPFKSNANIGYAOQHAH  
EFADLNLMEWQMRQEGDDEQVVRFGKVLFGQDDIKSVKVLSSGEGQRMILGK  
LMAKPNMLMDPPTNMDWESTESLNNALQYKTLFFVSHDRVFDLSLARIIEIR  
DGKITDFKGYSEYLSKRGVES"  
4834..7179  
/gene="VV1709"  
4834..7179  
/gene="VV1709"  
/note="identified by Genemark and Glimmer2"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC94473.1"  
/db\_xref="GI:37198639"  
/translation="WQKLRLVFCPTLVAPATFRQWYAAQTPAQVSAKALSTSDL  
SALPDLPMGPKEHIVNQLLSAVLELQTEGDAFALAOYKAPLAPOMVLNQQY  
LTLNLSGSKARLQDLTASRQLTGFPGVGTQFQRELATKINIELEVLQNSQY  
RALIKNTFISPVIVITGLKFAVYLLVMMANSKRIITLFPESQLBNINAPLMVK  
FIYVIRANKAIAIMLAIYTSALIASPLQHLVYLEIFTWMLIGSSAIAIFILEFA  
VYNSRSSKQVVALRELATIRRYVWSAIVAGVLIQSVRTIGKTIYYWVIALYSHFK  
LYVSYLRMKRKVPFVDRITERPVWOMALIKQDSPTLINPATIAVWVLIISHFK  
HRIATISNTWFSQALAYFRLEVAQKQSNQDNDSLYRIGDAPFOVLPPODSE  
WIDYASDELQNSRLTDSPALCVLSRGGVTSVLKAILGKVRNRIPLTNPIYD  
GYBELKQALATGGLAEESTEQILAFLKRSPTPIYIAVDNAQRLVKPVGQALQMK  
LTLNLSRQKNSHVVAIEKASVRFVDRAGRLFDWVSLPRNNKQISALISLRSV  
NODIENLSPDGLVVPKQMDSEETSEERAKQGFYLIWHYGDGNTVALRFRRLSLR  
RKESDSVVVRLFAPESEKLEQMPKMLAVRSIYQLEAVSEELSECTQSIASVI  
GTRFPEBSRGFTIMNEDKARVSDHMRITVTLDRQHLVAK"  
7181..8062  
/gene="VV1710"  
/transl\_table=1  
/product="small-conductance mechanosensitive channel"  
/protein\_id="BAC94474.1"  
/db\_xref="GI:37198640"  
/translation="MNKKVLAFLPAALISAPAFATEIANMENIQOPASIRNSGVFP  
SVIIATITWLLKFNNSMVESIGQFAQVWMLQKQSPFQFPFYTLAGLVPMFSR  
INDHIALIGGLTAVSGFALKDLAASRIGLTVMDIRPQVNDGRFTPEGNDITAI  
GLASVRRRTITDITITPNNKFLNEVTSVSNALMOVITPFYVWDEDTVALRELI  
QEAASSSRVTHLPRPVVAVKQTTNDNYALIQTLCAVYVDIAVEELPESDITLRVWK  
EFKKGKXPRPIASQ"  
complement (8150..10276)  
/gene="VV1711"  
complement (8150..10276)  
/gene="VV1711"  
/transl\_table=1  
/product="signal transduction histidine kinase"  
/protein\_id="BAC94475.1"  
/db\_xref="GI:37198641"  
/translation="MAINRRRHYLIAMHAKMSANTKSGKRRKIMQSDOTLWTRY  
RKSMVRYRLFTFSAPITFTLALIGISVYNSHTWSALIDVSRIGVANNSTLY  
LQKQAVYVAFABSTQFRTRINQPDQDQVHVKQKQYQDLDELAFRVDITBEK  
FRFLDLTKRSFPDVASETELQDLDPDLAPRAVPLIKTGKLTGRLISTVLPENQ  
SNDLIGLIDGLILNNSVTVADQIRDLIVIGDEDSRPIGTVLFMEDLSTVSNPLD  
SEDOIGRAITRVSGEYKSVLQOGERWVRAVVDAMVITAYQIRDOYNNITGLY  
TGYLMPFVKAVALTNIYEISVITLTLVSGLLVYGRSDLPFRIRIKHVVLYQDG  
KDKRIGLIGDENHELAQAKOPDNMDILKTRQOKOANLEKQVETASLNEK  
TEQLEHRIRLNTRDCLTNEKALAGETACIARHINPTAVIICNVLEIPELGE  
DAKAVEEIEFTVAQIDRLNITRSLQISQSGQVDEITWQVNPITSESTLYTGC  
SKRTTIEFTVELNHNSTSEVNRHQQLQILVNLQMANLHANNGRRLTISEDLEGE  
SVGAVIHIEPDQGGIKXPEKLSRIFDPEYTKBDGTGLGSQSLSSGSEKIVKVSQ  
LGISTSVYLPKRNKALRDTLLIG"  
complement (10173..11678)  
/gene="VV1712"  
complement (10173..11678)  
/gene="VV1712"  
/codon\_start=1  
/transl\_table=1  
/product="sigma-54 dependent response regulator"  
/protein\_id="BAC94476.1"  
/db\_xref="GI:37198642"  
/translation="MIHWTMSYTSFCELSMSIQDLPNSVSGYAPSVLYWVDEGM  
QALIKLAGWFSKVDVSGSIEQAEKURVNNHYDLVLDNLPRSGTIEBEAFNPE  
RKADVIFMTYADLETAITALKIGADTFLAKPNDLQMDQAVRCKDRILNARMQYAL  
RKDORNHISEMVGNSDKTQLKMITQFAPSKACVYLIEGSESGEVLVARGHLASG  
RTGPFVPLNCGALAPBELSEELFGHTSGAFTGAKKAREGLFVALIGTFLDEIGEMP  
LAWOSSILRVLEORSIRPVSSEKEIPVDVVAATNRDLOQEVKGSFRDOLYYRLNV  
LKTIEPPLRRKADQLQIVPFTKQSLGELMAGAPQAHEDITAMGDYSRGNVRELR  
NLIERCITLGGPRAHTWRELNGSRNLPNVSITVSHSAYPTFGGKKEPIGYGVPRNDW  
TLKVERAHAIQOVVHYDGNKSAASRDLGVARKTLEKRYKEMTEADYAE"  
11942..12760  
/gene="VV1713"  
11942..12760  
/gene="VV1713"  
/codon\_start=1  
/transl\_table=1  
/product="ABC-type tungstate transport system, permease  
component"  
/protein\_id="BAC94477.1"  
/db\_xref="GI:37198643"  
/translation="MNNVALTLAATSTTVSFSFSADVDYHRYALATTSTYSGULD  
YLPEPKDGVYVDVLAAGTGSIKKNGGVLDLVWTHAPKAEATVEKGYGLPJK  
LMVNDFTVNGPKADPAVDQDNSVLRYFKAIADNNVTFTSRGDSGTHKKEIGIWAOT  
KMEHNFQYHSVCGKQPTNLNMSNGTTLTRGTALANNNKDLAVIPEGOKLEN  
PYQVILNPERRISYINQAKAPSDWLNBRGGLINSFRLNGQLVVAADQX"  
12767..13465  
/gene="VV1714"  
12767..13465  
/gene="VV1714"  
/codon\_start=1

CDS  
7181..8062  
/gene="VV1710"  
/codon\_start=1  
/transl\_table=1  
/product="small-conductance mechanosensitive channel"  
/protein\_id="BAC94474.1"  
/db\_xref="GI:37198640"  
/translation="MNKKVLAFLPAALISAPAFATEIANMENIQOPASIRNSGVFP  
SVIIATITWLLKFNNSMVESIGQFAQVWMLQKQSPFQFPFYTLAGLVPMFSR  
INDHIALIGGLTAVSGFALKDLAASRIGLTVMDIRPQVNDGRFTPEGNDITAI  
GLASVRRRTITDITITPNNKFLNEVTSVSNALMOVITPFYVWDEDTVALRELI  
QEAASSSRVTHLPRPVVAVKQTTNDNYALIQTLCAVYVDIAVEELPESDITLRVWK  
EFKKGKXPRPIASQ"  
complement (8150..10276)  
/gene="VV1711"  
complement (8150..10276)  
/gene="VV1711"  
/transl\_table=1  
/product="signal transduction histidine kinase"  
/protein\_id="BAC94475.1"  
/db\_xref="GI:37198641"  
/translation="MAINRRRHYLIAMHAKMSANTKSGKRRKIMQSDOTLWTRY  
RKSMVRYRLFTFSAPITFTLALIGISVYNSHTWSALIDVSRIGVANNSTLY  
LQKQAVYVAFABSTQFRTRINQPDQDQVHVKQKQYQDLDELAFRVDITBEK  
FRFLDLTKRSFPDVASETELQDLDPDLAPRAVPLIKTGKLTGRLISTVLPENQ  
SNDLIGLIDGLILNNSVTVADQIRDLIVIGDEDSRPIGTVLFMEDLSTVSNPLD  
SEDOIGRAITRVSGEYKSVLQOGERWVRAVVDAMVITAYQIRDOYNNITGLY  
TGYLMPFVKAVALTNIYEISVITLTLVSGLLVYGRSDLPFRIRIKHVVLYQDG  
KDKRIGLIGDENHELAQAKOPDNMDILKTRQOKOANLEKQVETASLNEK  
TEQLEHRIRLNTRDCLTNEKALAGETACIARHINPTAVIICNVLEIPELGE  
DAKAVEEIEFTVAQIDRLNITRSLQISQSGQVDEITWQVNPITSESTLYTGC  
SKRTTIEFTVELNHNSTSEVNRHQQLQILVNLQMANLHANNGRRLTISEDLEGE  
SVGAVIHIEPDQGGIKXPEKLSRIFDPEYTKBDGTGLGSQSLSSGSEKIVKVSQ  
LGISTSVYLPKRNKALRDTLLIG"  
complement (10173..11678)  
/gene="VV1712"  
complement (10173..11678)  
/gene="VV1712"  
/codon\_start=1  
/transl\_table=1  
/product="sigma-54 dependent response regulator"  
/protein\_id="BAC94476.1"  
/db\_xref="GI:37198642"  
/translation="MIHWTMSYTSFCELSMSIQDLPNSVSGYAPSVLYWVDEGM  
QALIKLAGWFSKVDVSGSIEQAEKURVNNHYDLVLDNLPRSGTIEBEAFNPE  
RKADVIFMTYADLETAITALKIGADTFLAKPNDLQMDQAVRCKDRILNARMQYAL  
RKDORNHISEMVGNSDKTQLKMITQFAPSKACVYLIEGSESGEVLVARGHLASG  
RTGPFVPLNCGALAPBELSEELFGHTSGAFTGAKKAREGLFVALIGTFLDEIGEMP  
LAWOSSILRVLEORSIRPVSSEKEIPVDVVAATNRDLOQEVKGSFRDOLYYRLNV  
LKTIEPPLRRKADQLQIVPFTKQSLGELMAGAPQAHEDITAMGDYSRGNVRELR  
NLIERCITLGGPRAHTWRELNGSRNLPNVSITVSHSAYPTFGGKKEPIGYGVPRNDW  
TLKVERAHAIQOVVHYDGNKSAASRDLGVARKTLEKRYKEMTEADYAE"  
11942..12760  
/gene="VV1713"  
11942..12760  
/gene="VV1713"  
/codon\_start=1  
/transl\_table=1  
/product="ABC-type tungstate transport system, permease  
component"  
/protein\_id="BAC94477.1"  
/db\_xref="GI:37198643"  
/translation="MNNVALTLAATSTTVSFSFSADVDYHRYALATTSTYSGULD  
YLPEPKDGVYVDVLAAGTGSIKKNGGVLDLVWTHAPKAEATVEKGYGLPJK  
LMVNDFTVNGPKADPAVDQDNSVLRYFKAIADNNVTFTSRGDSGTHKKEIGIWAOT  
KMEHNFQYHSVCGKQPTNLNMSNGTTLTRGTALANNNKDLAVIPEGOKLEN  
PYQVILNPERRISYINQAKAPSDWLNBRGGLINSFRLNGQLVVAADQX"  
12767..13465  
/gene="VV1714"  
12767..13465  
/gene="VV1714"  
/codon\_start=1

gene

CDS

gene

CDS

gene

CDS

gene

CDS



gene  
CDS  
LELSANDKAYSFVAAKEPALHPGOSAAIVVDGKEIGVIGVHPBELERRKNGRTIV  
EIKMSIKNKVIPDEAVALSKEPPANRBDIAVVYDEVASGDYTNACLEPIGSEPLKAAKL  
PBYVGVGSEBEGKSLAIALTLQNSERTLEDADIAGADVALTVHSEKFGASLND"  
/complement (3584..4897)  
/locus\_tag="VV12372"  
/complement (3584..4897)  
/locus\_tag="VV12372"  
/note="COG1488"  
/codon\_start=1  
/transl\_table=11  
/product="Nicotinate phosphoribosyl transferase"  
/protein\_id="AA010746.1"  
/db\_xref="GI:27361840"  
/translation="MCAPI.FOBAPIQSVLD.LVYKINMGOAAYRPOTORYELIV  
SDNLSLDIVBEYREELNRLAERPDHQAQAIYAEKAPITAEFLSYLETFRHPQOY  
SVGIFRTAQDCQRLVINGIHETILVETLVMSITSELNRRKVAQ.PQSLHVLSE  
DQDLNLSLKRNTNFRSEMGRRRSFPAQKTMLDVLAARVPELLLTSTNHLA  
OENFLPTIGTVAHEMTMAQALVALQHSQVRLADKMLEAFNGSLGIALTDIGDALI  
SPDLDKATAYAGVRHDSGPSVWGDKI IAHESLGIPTTKLITPTGLDFARALDI  
CEYFAGRAQISFGIGFTLMDGMNTNDGTGPI SMVYKMAECNGSPVAKISDEPE  
KMCEDIFFLMLKORFGLVDLDAIETLKMKKQCKKRIQSV A"  
5176..5889  
/locus\_tag="VV12373"  
5176..5889  
/locus\_tag="VV12373"  
/note="COG1051"  
/codon\_start=1  
/transl\_table=11  
/product="ADP-ribose pyrophosphatase"  
/protein\_id="AA010747.1"  
/db\_xref="GI:27361841"  
/translation="MIVTIDMTCLKGEGKLEVLIKRTNPERPQGMMSI.PGGEVFE  
HNDISGGQPADPDSBARRI CROKITHYPIYISEPMVDGPKDIPWMSITLHYA  
LINTDIOIIONCG.CDOOLAMPALLESVUCEVALAPHADLIILAKTLAILEYTS  
VILFAEKELFVSDITHAYAKGVVDVNRRTIKRLIDGVIVSANKIKASTKKGKGR  
AQQYSLADKHVYFOTCLR"  
5902..6546  
/locus\_tag="VV12374"  
5902..6546  
/locus\_tag="VV12374"  
/note="similar to nicotinamide; COG1335"  
/codon\_start=1  
/transl\_table=11  
/product="Amidase"  
/protein\_id="AA010748.1"  
/db\_xref="GI:27361842"  
/translation="MYDIKPKRTASLDVDPQGFSELCNDEL.PVSGALIEYALKONH  
TRAKTLVSRDMEPRGAAWEATPRANKLEPVGLPEYDIKMPHCIVGQVILLPLP  
ALREYFQNMKGIDPDHPYGAETHDQADUTLSGGIEFLKARITDVLVGGIADPCV  
KXSIQLVADPFVILNLASTRAVPENNPVVSHLALGVICIDVNOQLHE"  
6962..7249  
/locus\_tag="VV12375"  
6962..7249  
/locus\_tag="VV12375"  
/note="COG0776"  
/codon\_start=1  
/transl\_table=11  
/product="Integration host factor, alpha subunit"  
/protein\_id="AA010749.1"  
/db\_xref="GI:27361843"  
/translation="MALTRADLAENLFEKLGFSKQDAQDTVEFEEIRKALNEBOY  
KLSGFNPLDRKNERPGRNPKTGEDIPTARAVVTFPRGQCLKARVENLK"  
complement (7347..8123)  
/locus\_tag="VV12376"  
complement (7347..8123)  
/locus\_tag="VV12376"  
/note="COG2357"  
/codon\_start=1  
/transl\_table=11  
/product="Conserved hypothetical protein"  
/protein\_id="AA010750.1"  
/db\_xref="GI:27361844"  
/translation="MSLFLRTTALMLVLVSRAPAPAIPTASNOEKORETSONEVCS

gene  
CDS  
KLFKHSISGLVGIQSDTQPLQPYTDPVLYKSAHQARELENTYKSTALDTPQYF  
AGVRSQGRAKQKIALENNOPEKITDLAKRTIVADVASVSAFTEALEETRTVKYK  
RFRSPGSEYDELALIVQLEKTNLVAEVLHKLALADVSQPEPHDLYQOI QWERRQAA  
SETRSDINVELATIKRLRSKSDLYOQAAQNPVLTTHLEA"  
8413..9066  
/locus\_tag="VV12377"  
8413..9066  
/locus\_tag="VV12377"  
/note="COG0500"  
/codon\_start=1  
/transl\_table=11  
/product="Thiopurine methyltransferase"  
/protein\_id="AA010751.1"  
/db\_xref="GI:27361845"  
/translation="MWDARWENKASNOIGFLEVDVNPDLRFRGALAPKEETLVY  
PLGKREDLAWLTKEDHVGASLSLIARSFAEHPVPTVPTISGQHELVQFDLS  
IVAGDEPTAPLSAEIDLYRMLIALPEMRVEYVIRIGLNPGRILLVSLDYPQO  
ENAGPPSVLYOEIEHFLFAGMYTRLQYODIADHHKRIAKOGLRSESEVYIENDK"  
complement (9119..10294)  
/locus\_tag="VV12378"  
complement (9119..10294)  
/locus\_tag="VV12378"  
/note="COG0027"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoribosylglycinamide formyl transferase 2"  
/protein\_id="AA010752.1"  
/db\_xref="GI:27361846"  
/translation="MFGIATRESATVLLIGSELEKEVAIECORIGLEVIACDRYP  
APMQVAHRSYVFDMLDASLEKVIADBPAPVPEIEAIDTKLELEBQGLNVPS  
AKATKLTNNKNEGRILVAEELGTTSPFADYQOFVAVEVAS.IPCVVPKSSG  
KQGSVKSIPADIEKAWQYAOEGRTAGVYIEGFDIPYEITLTPRAVDGVHPCAP  
IGHROEDVDRENOPOQMSENA IKAETVAQVNAALGGYGLFVGLFKVGDYFN  
EVSRRPHDQTLVTLISQMSFALHVAFTGMVKNITVOYGPASAVIIGNGOSELR  
PDGMSDALBPQQLRPFKPDINGRRRLGVVITRRSSTERKAVDAIIEBAKIKITV"  
complement (10426..11334)  
/locus\_tag="VV12379"  
complement (10426..11334)  
/locus\_tag="VV12379"  
/note="COG0295"  
/codon\_start=1  
/transl\_table=11  
/product="Cytidine deaminase"  
/protein\_id="AA010753.1"  
/db\_xref="GI:27361847"  
/translation="MQLQGFYTKSRFENALASAPESLSHAPILIIAADFASLSTVO  
FDHLKQTMNDLQVALLPFAAAYSAPISEFYGALVRLGALVGAWEFPGV  
OLGQTVARQSAISHAMMGEOGLSDITINPSCGHROFMELSSAKELKIQLPERE  
EKXLDHYLPDSFSPSLDGIESALMSGVHGFATFEDDLMQRAVEAMNSHAPYENL  
SGVALQTESKRVTLGAVAENAFNBSLPLOVALIOLLAGERFENIQSHALVESHKG  
KISHLACTQSTLEALNPDIPVSYLSL"  
complement (11675..11812)  
/locus\_tag="VV12380"  
complement (11675..11812)  
Query Match 42.5%; Score 339; DB 1; Length 306147;  
Best Local Similarity 66.5%; Pred. No. 1.7e-88;  
Matches 486; Conservative 0; Mismatches 245; Indels 0; Gaps 0;  
gene  
CDS  
184 AAATGGGCAATGCTGGCGGAGATAGTATGACGAAGATATGACCCGCTCATGCTGAGAA 243  
5275 ATGTGGCTATTCGGGAGAGATTTGTGTTGAACAAGATTATCCAAAGGAGAGACAA 5334  
244 CCTGTGANTAGAGATTTGATGAGAGGACGACGATTTGTTGGGAAAAAGTCCTACT 303

Db 5335 CCAGCTGACGCTGATTTGATTCCTGCGGACGCGATCTGTGTGGGAAAAATGATCC 5394  
 QY 304 TATCCCAATTTATATGCGCATCCGCTGGTGTATGCGAACCCCAACCGGATCCGATGCT 363  
 Db 5395 TATCCCTACCTACATGAGTACGCGGATGATGATGATGATGATGATGATGATGATGAT 5454  
 QY 364 TGGAGTGTGATGATTTTCCCATTAAGCTTATTAACCCCGTGAATGTCAACAAATAGAA 423  
 Db 5455 TGGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5514  
 QY 424 GATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483  
 Db 5515 AACTGT 5574  
 QY 484 GAATGT 543  
 Db 5575 GAATGT 5634  
 QY 544 GCTGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 603  
 Db 5635 GCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5694  
 QY 604 GATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 663  
 Db 5695 GATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5754  
 QY 664 CGTTGAT 723  
 Db 5755 CGTTGAT 5814  
 QY 724 AAAGAGCCAAACCAAGCCACCGTTATCGTCTGCGAGTCATGAATGCAACCTATTTTCA 783  
 Db 5815 AAGAGGAGCAAAACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5874  
 QY 784 ACCTGTTTACG 794  
 Db 5875 ACTTGCTTGG 5885

RESULT 4  
 AP005085/c 303450 bp DNA linear BCT 05-MAR-2003  
 LOCUS AP005085  
 DEFINITION Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 2/6.  
 ACCESSION AP005085 BA000032  
 VERSION AP005085.1 GI:28808465  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 . Vibrio parahaemolyticus  
 Vibrio parahaemolyticus  
 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 Vibrionaceae; Vibrio.  
 REFERENCE  
 AUTHORS  
 1 Nasu, H., Iida, T., Sugahara, T., Yamachi, Y., Park, K.S., Yokoyama, K.,  
 Makino, K., Shinagawa, H. and Honda, T.  
 TITLE A filamentous phage associated with recent pandemic Vibrio  
 parahaemolyticus O3:K6 strains  
 JOURNAL U. Clin. Microbiol. 38 (6), 2156-2161 (2000)  
 MEDLINE 20295086  
 PUBMED 10834969  
 REFERENCE  
 AUTHORS  
 2 Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,  
 Tagemori, K., Iijima, Y., Neijima, M., Nakano, M., Yamashita, A.,  
 Kubota, T., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H.,  
 Hattori, M. and Iida, T.  
 TITLE Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
 distinct from that of V. cholerae  
 JOURNAL Lancet 361 (9359), 743-749 (2003)  
 MEDLINE 22508454  
 PUBMED 12620739  
 REFERENCE  
 AUTHORS  
 3 Ibaes 1 to 303450)  
 Oshima, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T.,  
 Honda, T., Shinagawa, H., Hattori, M. and Iida, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome

INFORMATION  
 COMMENT  
 This clone was isolated from a patient presenting with acute gastroenteritis.  
 FEATURES  
 SOURCE  
 Location/Qualifiers  
 1. 303450  
 /organism="Vibrio parahaemolyticus"  
 /mol\_type="genomic DNA"  
 /strain="O3:K6"  
 /sub\_strain="RIMD 2210633"  
 /db\_xref="taxon:670"  
 /chromosome="2"  
 81..1574  
 /gene="VPA0300"  
 81..1574  
 /gene="VPA0300"  
 /note="similar to EMB:CA6187.1 (AL591787) percent identity 68 in 487 aa"  
 /codon\_start=1  
 /evidence=not experimental  
 /transl\_table=11  
 /product="hypothetical signal peptide protein"  
 /protein\_id="BAC61643.1"  
 /db\_xref="GI:28808465"  
 /translation="MKLVKPKRLVLTLLASLPISVAYAPNANESSIVQOQWGNWPTQNEKRVQELVYQRAISAYQAPNANVIGMDSSSKFGKINVLPIKMDKSRILVPTPNADVVMYVLDLKEKTPVVAAPKVGMTDFQRTITVGLIGPRA RGVLYLLPENVDPVPKGFTFTSTNNVFLFRTVMKGGADSPVAKIADQRI IYPLMBEKKIOPMKFPNAGORINMVPDFSEYMKQFVDEPISAISSPELRVLI ASIGIVGKPEFIVSEKOLSEKAVTAPMILQRLQSKDEBNLYKQROVERAMA GATSEVMOSEYVLDIQRRAAFQYVASSAPAMVWRTNAGSKYPTTDRSDGIDLNGSH QYKHLIPANPACLPANVTLVNVDTMTPTSLLSKNGFMDIENXTDSDIDLVTGCP QLPNGAPESNTIKITIPGRDPLTALRLYGTGIEFFDQWPKDDVLTKI"  
 /complement(1835..2314)  
 /gene="VPA0301"  
 /note="similar to DBJ:BA843758.1 (AP003138) percent identity 57 in 158 aa"  
 /codon\_start=1  
 /evidence=not experimental  
 /transl\_table=11  
 /product="putative peptide methionine sulfoxide reductase"  
 /protein\_id="BAC61644.1"  
 /db\_xref="GI:28808467"  
 /translation="MKETKYPAGCGIWMGYQEFKRLHPVISTEAGRANKTDNTQSEY DGYACVRTPEFDSQVSIPLMGVFPFIIIDPSINKQGDVGEKRYTGVYSRUTKALE IAKSFATRTDADKIAVEVLPKNVVPDSDEHODRLTRFPNDYCHPLDLHKYXNK" 2493..2882  
 /gene="VPA0302"  
 2493..2882  
 /note="similar to GB:AA64344.1 (L16865) percent identity 52 in 129 aa"  
 /codon\_start=1  
 /evidence=not experimental  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC61645.1"  
 /db\_xref="GI:28808468"  
 /translation="MKIEHVAIWTERLEELKGFYKYPFNAVSNDRKYNPKFSSYFL SPESGARLEIMSWEGVTTCENSHMQVTGLAHAFALGSEQAVDQITKTLVBDGYORI DGPRTYGDGYBSCLDPDGNRIELTV" 2974..3714  
 /gene="VPA0303"  
 2974..3714  
 /note="similar to GB:AA120622.1 (AE008775) percent identity 54 in 243 aa"  
 /codon\_start=1



```
/evidence=not experimental
/transl_table=11
/product="putative regulatory protein"
/protein_id="BAC61646.1"
/db_xref="GI:28808469"
/translation="MNPRQWEILQVNDKRRQVTELSDIIGSVTLIEQDINFLPEQ
GYKRVGAATALQSDIDITRLERFDIKQTLANKADLVAREVTLIGGSAKALLA
RTLAERGDTITTPSAVIATLINTSANTIIIGVYQHOGESLVGPTLCLCENHFS
TATLGDGFRQDGTFRSDMRADIALAIKARRIVITDSKRGQIIPSSIGKINE
ISVLTDKAAPSDEQLKGLGVVVLG"
/complement(3733..4089)
/gene="VPA0304"
/complement(3733..4089)
/notes="similar to GB:AAF96321.1 (AE004376) percent
identity 76 in 117 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC61647.1"
/db_xref="GI:28808470"
/translation="MQHEKNTYERGTENIGATKAPFHVFGMOFVDYGDYAAPSG
QGLDGGFYSAEQVQSNGAALLVYSSDIHATLEKVAFGQIIRPIEPFGGRFH
FTSPSGNEPVMSESH"
/4319..5860
/gene="VPA0305"
/4319..5860
/gene="VPA0305"
/4319..5860
/notes="similar to REF:NP_231225.1 (NC_002505) percent
identity 83 in 510 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="catalase"
/protein_id="BAC61648.1"
/db_xref="GI:28808471"
/translation="MNSTHWRATQMSKSFLLITVGLASTLSQAQTLTRDNGAPVGD
ONSITAGENSVLQDVHLIQKQFARERIPRVVHAHTGAHGFVAGSDPDLTV
SAPTEKGVTPVFAFSTVTHSGSEPLRDPGFATGEYEGWMDVGNLNVFV
IRDSIKPRPMVHSIKSPVNTODPNRPFDPFSEBPAATHMLTWYSNIFPGASRTM
DGRVHAAYKKNQGDVNYTKFPMKSCQCKIKLRPKVTEMGKDPNHLTNLYAIE
KGVYFKDLTVKVLSPALSKDYNGLDITKWLNVDRKVGMTLNLNPEFLIETE
QSAFAPSNLIPGLEPSEDRLLQGRFVADTOLYRLGANLFDLPVNRPLTSVNNRNQ
GLSNNAQLSDNVNPSRKILAEQNOFPAVETKLVGVQOKAISPEDFQOAGVLY
RSNMEQDRSLINLAGLKNKVIDKDIKATWVSFYRADKEYGSLAEXATDNLQVY
NKAMM"
/5962..6429
/gene="VPA0306"
/5962..6429
/notes="similar to GB:AAF94738.1 (AE004235) percent
identity 63 in 152 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="ANKB protein"
/protein_id="BAC61649.1"
/db_xref="GI:28808472"
/translation="MKQVFLSSVLSFSLGSLFSLPASODKLETOYOSIVLELFFDAARI
GNNEVETPVSQGFIDORNNESYALMAAYOGKNDYTRLLDSGANA.CDDKQNT
ALMGALIKRISTAKDLYOAECSPLDKRNRAGDLKTFATYQGSVLSLQK"
/complement(6548..7201)
/gene="VPA0307"
/complement(6548..7201)
/notes="similar to GB:AAF42465.1 (AE002563) percent
identity 47 in 200 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative pyrazinamidase/nicotinamidase"
/protein_id="BAC61650.1"
```

```
/db_xref="GI:28808473"
/translation="MTYRVNVQCTTAVIDVPEKGSFSLCGDELVPAAGALVPEILKN
HAKRILVSRDLHPKPAANDAEITPANNKEPVGLPNDVKNRHCVTGTTGVLSDBL
EPVLVDYQVKNKGDPAHPYGIFFHVDATKKTGAENFKNKIDIVVGGALDLC
VKSVQWALDGFYFVIVNLAIRAVLPDITDVSVAEMKEKGLFVNKADIIVERFA"
/complement(7234..7959)
/gene="VPA0308"
/complement(7234..7959)
/notes="similar to GB:AAF96011.1 (AE004352) percent
identity 60 in 241 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC61651.1"
/db_xref="GI:28808474"
/translation="MIVTVDIIPRLSGCADKGLLEVLLIKRSPNRPYHGVWALPGF
VFDDLTSEGRPADENFEAARRICREKHTYPRHSEAFIDDPREDDESLNT
HYALVDNVEQINAGVPGCOLKMFLOALINGEETLAFDHOKEIKAWOKRASIE
YTVLILFALMKEFIVADITSAOYFSGIDISRMITKRRLISGVLKPTNKVASTKRG
GKPMVTVTLSDVTFPQNCIRG"
Query Match 35.4%; Score 282.8; DB 1; Length 303450;
Best Local Similarity 63.6%; Pred. No. 6.9e-72;
Matches 431; Conservative 0; Mismatches 247; Indels 0; Gaps 0;
120 TGGCGCGAATCTATCCAGGTTTACTGCGTGAAGACGCTTAATCCAAATCGGACGATG 179
7915 TCCGATTAAGAGATTGGAAGGTTCGATTAAACGTTCAATCGAATCGGCTTACCA 7856
180 TGGTAATGGCATTGCGTGGCGGAGATGATGATGACGAAGATGACCGCTCATGGTGG 239
7855 TGGTGTCGCGCATTTACCGGCGGCGGTTCCTTGAATGAAGCTTACGACGAGCGG 7796
240 AGAAGCTGTCATGAGATTTTGAATGACGAGACGACGATTTGTGCGCAAAAGTCCA 299
7795 AGCTCGCGGATGAAAACCTTTGAAGCGGCGCGCGACGCTTTGCGCGGAAGATTCA 7736
300 TACTTATCCATATTTTATCAAGCATCCGCTGTTGATGCGCAACCCAAACGCGATCCGA 359
7735 TACCTACCTGCTGACTTACGCGAAGGTTTATGACGAGAACCTTAAGCGCATCCGGA 7676
360 TGGTTGAGTGCATGATTTTCCCATTCACGCTTTTAAACCCGTGGAATGTCAAACAAT 419
7675 AGATTGAGCTTAAACATCCCTTACGATTTAGTTGACCGCAACACGCGAGCAAT 7616
420 AGAAGATTGGTATCGACCCCGAGCGCGCAATTTGGTTGATCTTCACTTACTCAA 479
7615 CAAATGCTGCGCTGCCAGATGTCAGCTCAATGTTCCCATTCGAGCGGATCTCAA 7556
480 AGAAGAAATGCGGTGCTTTGATCATGTCGCGCAATTCAGATGCGTGGCAAAATT 539
7555 TGAAGAAACCTTAGATTGACCAACAAAACCATGCAAAAACATGGCGAAACT 7496
540 ACGCGCTGCGGTGAATACATCCGCTGATCTATTCTTAGAAAAGGTTTATG 599
7495 GCGTCTTCATGATGATACATAGTGCCTGCTGCTGCACTCATTAAGATTTTGGT 7436
600 GCGGATTTATGATGCTTACGCCAATTTGGCGTGAAGTTAATGCAATGACCATTA 659
7435 CGCAGACATCATTTCTGATATCAAGATTTGGCATGACATGACGCGCATGACATTA 7376
660 ACGCGCTTATCATATACCGGGGATGATCGTCAAGTAAATTAATGGCCGATTTGTA 719
7375 GCGCGTTTATCATCTAGGTGATGAGCCGACCAACAAATCGGTGATCAACAA 7316
720 AGGCAAGAGCCAAACAGGACCGATTATGCTTTGCCAGTCAAGTCAACCTATT 779
7315 AGGCAAGCGGTAAAGCAGCATGATGTAACGCTGACCAAGCATGAAGTGAACCTTTT 7256
780 TCAAACTGTTTACGAGG 797
```

Db	7255	TCAAAACGTGTTGGCTGG	7238
RESULT 5			
LOCUS	AX083744	1141 bp	linear
DEFINITION	Sequence 22 from Patent WO0111061.		
ACCESSION	AX083744		
VERSION	AX083744.1	GI:1185472	
KEYWORDS			
SOURCE			
ORGANISM	Synthetic construct		
REFERENCE	1	artificial sequences.	
AUTHORS	Kunst, L. and Clemens, S.		
TITLE	Regulation of embryonic transcription in plants		
JOURNAL	Patent: WO 011061-A 22 15-FEB-2001;		
FEATURES	UNIVERSITY OF BRITISH COLUMBIA (CA)		
source	Location/Qualifiers		
	1..1141		
	/organism="synthetic construct"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32630"		
	1..1141		
	/note="consensus sequence of A.t., L.a., and B.n. PAB1		
	promoters"		
ORIGIN			
Query Match	5.3%; Score 42; DB 6; Length 1141;		
Best Local Similarity	8.3%; Pred. No. 0.62; Mismatches 357; Indels 3; Gaps 1;		
Matches	59; Conservative 295; Mismatches 357; Indels 3; Gaps 1;		
Qy	8 TCACAAAGCTTATTTACATTTGACCATCATGACCTAAGCTAAATGAAGCCGCTGT	67	
Db	YCYRRWYNNKSRWKKWYKKKYYBCAANTSBRYHARRKDKAYBMTMTNNKMGTGR	124	
Qy	68 TTATTTAGGAAGCACTTTGATGTGACATACGATATATTTGCTGCGTCTGGCCGA	127	
Db	125 HRWYRMRMBJVDHVVYTAANNAAVTTCCMDKDKRTRWKKNNATGMDDDTKYMW	184	
Qy	128 AATCATCAGATTTTACTGAGCAACGCTCTAATCCAAATCGGCGATGTGGTAAT	187	
Db	185 NNNGBVTWYVRYKTDWBSBKNNYGBMWKNNNSYVTVYWWDMCKRKYRWRVT	244	
Qy	188 GGGCATTCGCTGGCGGAGTAGTGTATGACGAAGATATACCGCTCATGCTGGAGAACTG	247	
Db	245 RGRWNTVYVAEBTARRRNNGWBBAMVRRWYNNNNNNNAKACRKYWGNRABVNS	304	
Qy	248 TCGATGAGATTTTGGATGACGAGAGACACGATATTTGTCGGCAAAAAGTCATCTAATC	307	
Db	305 TCTTW--KSKTTKVRISCAANNRCAGANADHKMKBSAAMGYVNNNNNNNTTYKAR	361	
Qy	308 CTAATTATTCAGCGATCGCTGCTGTATGCGCAACCCCAACGCGATCGAATGGTGA	367	
Db	362 HBAAWMDWYHSAKKWHAANAAYSRKKTWBYRKTMYVNNNGTTWKKMYAMWKMMDW	421	
Qy	368 GTGTCAGATTTCCATAGACCTTTATTAACCCGTGGAANGTCAAAAGAAATAGAAATT	427	
Db	422 BGTNNNNNGRATTYYGWNKKMMTYTKKANNCKRMRMDHKTCHNNNTTYKKKTYNN	481	
Qy	428 TTGGTATCGACCCCGAGCGCGCTAATGGTTGATCTTTCATACCTTACTCAAGAGA	487	
Db	482 CYWISMNTGKSHRBAALAYVTWYMMRRYAHANNNDWYMKACATYKCYB/CSKMNMYA	541	
Qy	488 TGCCGCTGCTTTGATGATGCGGCAAAATTCAGATGCGTGGCAAAAATTAAGCCGCTG	547	
Db	542 AMYKSSWNTYSRYRMRKTNNSMKRSDTRMBGRANNYARABHYGYKNTFRWMSHTWB	601	
Qy	548 CGGTGATACACATCCGCTGCTATTTATTTAGAAAAGAGTTTATGCGCGGATA	607	
Db	602 HBRGAALHYMMBYBAKCHCKMAYKAKYKAGAGSNNNNNNNNNNNNNNNNNNATTCARD	661	
Qy	608 TTATTTGATGCTACGCCAAATTTGGCTGCGAAGTTATGCGATACCAATTAAGCCGCT	667	

Db	Qy	Db	RESULT 6
662	YIIASRWTAMAAKRYIKYKAAANNAIYTHANNWGCANNADPTDTRITMKANNNNNAAGTWK	662	YIIASRWTAMAAKRYIKYKAAANNAIYTHANNWGCANNADPTDTRITMKANNNNNAAGTWK
722	NNNNNNNAKNAASAAKXVAAAAVKAKKKEWRMNKWAAGMHADAAABTTDKENNG	722	NNNNNNNAKNAASAAKXVAAAAVKAKKKEWRMNKWAAGMHADAAABTTDKENNG
LOCUS	AC138591	LOCUS	AC138591
DEFINITION	Mus musculus chromosome 3 clone RP24-224A19 map 3, WORKING DRAFT	DEFINITION	Mus musculus chromosome 3 clone RP24-224A19 map 3, WORKING DRAFT
ACCESSION	AC138591	ACCESSION	AC138591
VERSION	AC138591.3	VERSION	AC138591.3
KEYWORDS	HTG, HTGS, PHASE1, HTGS, DRAFT.	KEYWORDS	HTG, HTGS, PHASE1, HTGS, DRAFT.
ORGANISM	Mus musculus (house mouse)	ORGANISM	Mus musculus (house mouse)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 146963)	AUTHORS	1 (bases 1 to 146963)
JOURNAL	Unpublished	JOURNAL	Unpublished
AUTHORS	2 (bases 1 to 146963)	AUTHORS	2 (bases 1 to 146963)
TITLE	Barren, B., Nusbaum, C. and Lander, E.	TITLE	Barren, B., Nusbaum, C. and Lander, E.
JOURNAL	Mus musculus chromosome 3, clone RP24-224A19	JOURNAL	Mus musculus chromosome 3, clone RP24-224A19
AUTHORS	Unpublished	AUTHORS	Unpublished
REFERENCE	2 (bases 1 to 146963)	REFERENCE	2 (bases 1 to 146963)
AUTHORS	Barren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barne, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazaro, B., Chopel, Y., Collymore, A., Canarata, J., Chang, J., Chazaro, B., Chopel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dwar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gargyala, S., Gird, S., Graham, L., Grand-Pierre, N., Hafez, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Medlirm, J., Meneus, L., Mihova, T., Mieng, Y., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.	REFERENCE	2 (bases 1 to 146963)
JOURNAL	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	JOURNAL	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS	3 (bases 1 to 146963)	AUTHORS	3 (bases 1 to 146963)
TITLE	Barren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N., Anderson, S., Archchi, H.M., Barne, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chopel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fato, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gargyala, S., Graham, L., Grand-Pierre, N., Hafez, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Moblitt, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Medlirm, J., Meneus, L., Mihova, T., Mieng, Y., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.	TITLE	Barren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N., Anderson, S., Archchi, H.M., Barne, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chopel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fato, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gargyala, S., Graham, L., Grand-Pierre, N., Hafez, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Moblitt, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Medlirm, J., Meneus, L., Mihova, T., Mieng, Y., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
JOURNAL	Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	JOURNAL	Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Mar 23, 2003 this sequence version replaced gi:28275009. All repeats were identified using RepeatMasker. Smt, A.F.A. & Green, P. (1996-1997)	COMMENT	On Mar 23, 2003 this sequence version replaced gi:28275009. All repeats were identified using RepeatMasker. Smt, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

----- Project Information  
Contact: sequence\_submissions@genome.wi.mit.edu

Center project name: 128955

Center clone name: 224\_A.19

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 145843 bases at least Q40

Consensus quality: 146107 bases at least Q30

Consensus quality: 146232 bases at least Q20

Insert size: 146000; agarose-fp

Insert size: 146363; sum-of-contigs

Quality coverage: 15.5 in Q20 bases; agarose-fp

Quality coverage: 15.4 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2163: contig of 2163 bp in length  
\* 2164 2263: gap of 100 bp  
\* 2264 5396: contig of 3133 bp in length  
\* 5397 5496: gap of 100 bp  
\* 5497 11945: contig of 6449 bp in length  
\* 11946 12045: gap of 100 bp  
\* 12046 13688: contig of 763 bp in length  
\* 13689 19788: gap of 100 bp  
\* 19789 39404: contig of 19616 bp in length  
\* 39405 39504: gap of 100 bp  
\* 39505 140276: contig of 100772 bp in length  
\* 140277 140376: gap of 100 bp  
\* 140377 146963: contig of 6587 bp in length.

## FEATURES

## source

1. 146963  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="3"  
/map="3"  
/clone="RP24-224A19"  
/clone\_lib="RP24 Male Mouse BAC"  
1. 2163  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
2264. 5396  
/note="assembly\_fragment"  
5497. 11945  
/note="assembly\_fragment"  
12046. 19688  
/note="assembly\_fragment"  
19789. 39404  
/note="assembly\_fragment"  
39505. 140276  
/note="assembly\_fragment"  
140377. 146963  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

## ORIGIN

Query Match 5.0%; Score 40.2; DB 2; Length 146963;  
Best Local Similarity 52.7%; Pred. No. 3.5;

Matches 87; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 411 CAACCAATGAGATTGTTGTCATCCCGAGCGCGTAATGTTGATCTTATAC 470

Db 123522 CAGATGAAGAGAGATTGTTGCGCTATCCAACTGTCATCTTAAGAAACATA 123561

QY 471 TTACTCAAGAGAAATGCGCGCTGCTTTGATCATGTCGCGCAATTCAGATCGCG 530

Db 123582 ATTGCTAATTAGAGAGAGAGAGGCTTTCTCTGTGTCAGAGAAATTAAGTCGA 123641

QY 531 GCMAAATTCGCGCTGCGTTAATACATCCGTCATCTT 575

Db 123642 GAGAAATTCACACCTCATAGGAATATCATTCATTCGAGATT 123686

## RESULT 7

## LOCUS

AF146758 2067 bp DNA linear INV 09-JUN-2003

DEFINITION Ceratitis capitata putative amylase-related protein (Amyrel) gene,  
complete cds.

ACCESSION AF146758 GI:27447662

## VERSION

AF146758.1 GI:27447662

## KEYWORDS

Ceratitis capitata (Mediterranean fruit fly)

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

## AUTHORS

## TITLE

## JOURNALS

QY 512 CCCAATTCAGATGCGTGGCAAAATTAACGGCTGCGGTGAATACATCCGTGGAC 571  
 DB 1879 GAGAGGTTTTCGTAATCAAGTAAGACAGCTGAGCTTCAGAGAAATGCGTACTCT 1938  
 QY 572 TATTTTCATTAGAAAAGATTGTTAGTGGCGAATATTATGATCCCTAGCCCAA 627  
 DB 1939 TTTTATTAAGAAAATTTTATGCGGTGCTTATGCTTTGACACCAAA 1994

RESULT 8  
 AC012853/c 16237 bp DNA linear HTG 03-NOV-1999  
 LOCUS AC012853  
 DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 ACCESSION AC012853  
 VERSION AC012853.1 GI:6223468  
 KEYWORDS HTG; HTGS PHASE2.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 16237)  
 AUTHORS Adams, M. and Venter, J.C.  
 REFERENCE Direct Submission  
 TITLE Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
 COMMENT This sequence was identified as CDM:10210833 by the submitter. For further information on this sequence you may e-mail to fly@celera.com.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 Location/Qualifiers  
 1..16237  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"

ORIGIN  
 Query Match 4.9%; Score 39.2; DB 2; Length 16237;  
 Best Local Similarity 47.5%; Pred. No. 5.5;  
 Matches 149; Conservative 0; Mismatches 163; Indels 2; Gaps 1;

QY 330 GGTGATGCGCAACCCCAAGCGATCCGATGTTGAGTCAATTTCCCATTCGC 389  
 DB 2498 GGTAAATTAATTCATTAACCGGAATGATGGGTAATCTAAATATTAATGTTAATCAT 2439  
 QY 390 TTTATTAACCCGTGATGTCACAAATAGAAATTTTGTATTCACACCCGCGCGC 449  
 DB 2438 ACTTACTACTGTTTAAATGCAAGTAACGAAATTTTAAATTAATTAATTAATGTC 2379  
 QY 450 TAAATGTTTGAATCTTCACTTACTTCAAGAAAGAAATGCGGCTTGTGATCATGT 509  
 DB 2378 TAAATGATATATATCTTTTAAACGCCAGAAAGAAAGCTATATCAGTTCAATATGT 2319  
 QY 510 CGCGCAATTCAGCATGCGTGGCAAAATTAACGGCTGCGGTGAATACACATCGGTGT 569  
 DB 2318 GAATCAAAACAGGCAATAAACTGAAATCCAGCAATATCTACCTTAACAGATCCCTATT 2259  
 QY 570 ACTATTTTCATTA--GAAAAAGATTTTATGCGGGAATATTAATGATGCTACGCCAA 627  
 DB 2258 CACATGCAATGTATGAATTAATTTGTAGACGGCAATATTTGCAATATCAACGTAGA 2199  
 QY 628 TTTGCGCTGGAAGT 641  
 DB 2198 TTCGACATCGGACT 2185

RESULT 9  
 AC012650/c 74085 bp DNA linear INV 12-MAY-2001  
 LOCUS AC012650

DEFINITION Drosophila melanogaster, chromosome X, region 19B-19E, BAC clone BACN24N09, complete sequence.  
 ACCESSION AC012650  
 VERSION AC012650.8 GI:14028960  
 KEYWORDS HTG.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 74085)  
 AUTHORS Celisner, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amaratunga, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Bantz, J., Beeson, K.V., Busan, D.A., Carlson, J.W., Center, A., Change, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R.F., Gary, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., Mincosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nuno, J., Paclob, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phoumenavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svitzkas, R., Tector, C., Williams, S.M., Zaveri, V.S., Smith, H.O., Rubin, G.M. and Venter, J.C.  
 REFERENCE Sequencing of Drosophila chromosome X, region 19E-19E  
 2 (bases 1 to 74085)  
 TITLE Celisner, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G., Butenoff, C., Change, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Katta, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomocan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclob, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snit, E., Svitzkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.  
 JOURNAL Direct Submission  
 COMMENT Submitted (02-NOV-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On May 12, 2001 this sequence version replaced gi:5958011.  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgs@fruitfly.berkeley.edu](mailto:bdgs@fruitfly.berkeley.edu).  
 Location/Qualifiers  
 1..74085  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /strain="hy; cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="X"  
 /map="19E-19E"  
 /clone="BACN24N09 (D1133)"  
 /clone\_11b="RP1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBelBac11)"

ORIGIN  
 Query Match 4.9%; Score 39.2; DB 3; Length 74085;  
 Best Local Similarity 47.5%; Pred. No. 6.5;  
 Matches 149; Conservative 0; Mismatches 163; Indels 2; Gaps 1;

QY 330 GGTGATGCGCAACCCCAAGCGATCCGATGTTGAGTCAATTTCCCATTCGC 389  
 DB 67158 GGTAAATTAATTCATTAACCGGAATGATGGGTAATCTAAATATTAATGTTAATCAT 67099  
 QY 390 TTTATTAACCCGTGATGTCACAAATAGAAATTTTGTATTCGACCCGAGCGCGC 449

Db	67098	ACTTACTTACTGTTTATTTGCAAGTAACTCGAAATTTAAGTGATTCAAATATTTGGC	67039
Qy	450	TAATTGGTTATCTTCACTACTTACTCAAAAGAAATTCGGCTGGCTTTGATCATGT	509
Db	67038	TAAATAGTATATATCTTTTAAAAAGCCACGAAAGAAAGCCATATTCAGTTCTAAATGT	66979
Qy	510	CGCGCAATTTCAGCATGGGTGGCAAATTTACGGCGTCGGTTGAATACATCCGTGT	569
Db	66978	GAATCAAAACAGGATATAAACTGAATACACAGCCAACTACCGTAAACGATTTCTAAT	66919
Qy	570	ACTATTTTCATTA--GAAGAGATTTTATGGCGGATATTAATTTGATGCTTACGGCAA	627
Db	66918	CACATGGGAATGTATGATATTAATTTGTAGACGGCGAATTTGTGCATATATCAAGTAA	66855
Qy	628	TTTGGCGTCAAGT	641
Db	66858	TTGCGACTCGGACT	66845

LOCUS	DEFINITION	AC011761	108350 bp	DNA	linear	INV 06-SEP-2001
AC011761	<i>Proscophila melanogaster</i> , chromosome X, region 19F-19E, BAC clone					
BACH50605	<i>Proscophila melanogaster</i> , complete sequence.					

ACCESSION	AC011761	
VERSION	AC011761.14	GI:15451480
KEYWORDS	HTG.	
SOURCE	Drosophila melanogaster (fruit fly)	
ORGANISM	Drosophila melanogaster	

ORGANISM: Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephraeroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS

Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Brandon, R. C., Rogers, Y. Y., An, H., Baldwin, D., Bonzon, J., Beeson, K. Y., Buetan, D. A., Carlson, J. W., Center, A., Chame, M., Davenport, L. B., Dietz, S. M., Dodson, K., Dorsett, V., Doup, L. E., Doyle, C., Dreeneh, D., Farrow, D., Ferrante, S., Frishe, E., Gault, R. F., Garg, N. S., George, R. A., Gonzalez, M., Houch, C., Hoskins, R. A., Hostin, D., Howland, T. J., Ibegwam, C., Jalali, M., Kruse, E., Lall, P., Mattei, B., Moshirif, A., McIntosh, T. C., Moy, M., Murphy, B., Nelson, C., Nelson, K. A., Nunn, J., Pacich, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phoonanavong, S., Piltman, G. S., Piri, V., Richards, S., Scheele, F., Stabile, C. M., Strong, R., Strydom, R., Tector, C., Williams, S. M., Zaveri, J. S., Smith, H. O., Rubin, G. M., and Venter, J. C.

Sequencing of Drosophila chromosome X, region 19E-19E

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Unpublished  
2 (bases 1 to 108350)  
Celisner, S. E., Adbayani, A., Arcaina, T. T., Baxter, E., Blazek, R. G.,

Benhoff, C., Champagne, M., Chaver, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummachi, S.R., Katra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomcan, M.A., Mazza, P., Moinirei, A.R., Moshrefi, M., Nixon, K., Paelele, J.M., Park, S., Pfeiffer, B., Poon, L., Segueira, A., Seidl, H., Sitr, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zheng, R., Zieran, L.L. and Rubin, G.M.

**Title** Direct Submission

**Journal** Submitted (14-OCT-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

**Comment** On Sep 6, 2001 this sequence version replaced gi:14280137.

Sequence submitted by: Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgsp@fruitfly.berkeley.edu](mailto:bdgsp@fruitfly.berkeley.edu).

**FEATURES**  
**Source**

```

FEATURES
    source
        Location/Qualifiers
            1..108350
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /strain="Y; cn bw sp"
                /db_xref="taxon:7227"
                /chromosome="X"
                /map="19E-155"
                /clone="BACHS0G05 (D134)"
                /clone_1b="PPCI-98 (Roswell Park Cancer Institute
                Drosophila melanogaster BAC library, partial EcoRI in
                pBelBAC11)"

```

Query Match	4.9%;	Score 39.2;	DB 3;	Length 108350;
Best Local Similarity	47.5%;	Pred. No. 6.8;		
Matches 149;	Conservative 0;	Mismatches 163;	Indels 2;	Gaps 1

Db 104731 GGTAAATTATCATTAACGCGAATTGAATGGTGATCTAAATATTACTGTTACAT 104790

Db 104731 ACTTACTACTTGTTAATTGCAAGTAACGTGAATTTAAGTATTCAATAAATTGGC 104856

Db 104851 TAAATAGTATATATCTTTTAAACGCCAGAAAGAAAGCCTATATCACTTCTAAATGT 104910

Db 104911 GAATCAAAACAGCGCATAAAACTGAAATACAGCCATATCTACCGTAAACAGATTCTATT 104970

DB 104971 CACATGCGAATGTATGAAATAATAATTGTAGACGGCGGAATTTGTGCATAATCAACGTAGA 105030

QY 628 TTTCGGCTCGAAGT 641

DD I05031 IUGSAC:IUGSAC I 050044

LOCUS	AE003571	302540 bp	DNA	linear	INV 14-FEB-2003
DEFINITION	Drosophila melanogaster chromosome X section 68 of 74 of the complete sequence.				

VERSION	AE0035/1.3
KEYWORDS	G1:22832655
SOURCE	Drosophila melanogaster (fruit fly)

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

Amanatides, P.G., Schreier, S.B., Li, P.W., Robbins, R.A., Galle, R.C., George, R.A., Lewis, S.B., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

Abril, J. F., Agbeyani, A., An, H. T., Andrews-Pfankoch, C., Baldwin, D., Baller, R. M., Babu, A., Baxendale, J., Bayraktarov, L., Beasley, E. M., Beeson, K. V., Benson, P. V., Bernier, R. P., Bhattacharya, D., Bolshakov, S.

Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D.,  
Wan, R.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabot, G.L.,  
April, J.F., Abbayanti, A., An, H.J., Andrews-Pfannkuch, C., Baldwin, D.,  
Ballaw, R.M., Bastru, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,  
Beeson, K.Y., Benos, P.V., Bernam, B.F., Bhandari, D., Bolshakov, S.,  
Borkova, D., Botchan, M.R., Bouck, T., Brockstein, P., Brottier, P.,  
Burtis, K.C., Bussan, D.A., Butler, H., Cadieu, E., Center, A.,  
Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Daveport, L.B.,  
Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I.,  
Diets, S.M., Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S.,  
Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C.,  
Fertner, S., Fleischmann, W., Foster, C., Gabriellian, A.E., Gary, N.S.,

Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpén,G.H., Ke,Z., Kenton,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Krat,C., Kravetz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Ma,C., Matel,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Mikhina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Murray,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusken,D.R., Pacleb,J.M., Palazolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reiner,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Klamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svitzkas,R., Tector,C., Turner,R., Venter,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A., Weinstein,G.M., Weisenbach,J., Williams,S.M., Woodgett, Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,U., Yen,R.F., Zaveri,J.S., Zhang,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of *Drosophila melanogaster*  
 Science 287 (5461), 2185-2195 (2000)

20196006  
 10731132

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

2 (bases 1 to 302540)  
 Celinker,S.E., Adams,M.D., Kronmiller,B., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., Banson,J., An,H., Baldwin,D., Bazon,J., Beeson,K.Y., Busan,D.A., Carlson,J.W., Center,A., Chape,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorssett,V., Doup,L.E., Doyle,C., Dresnek,D., Fattán,D., Fertler,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A., Gonzalez,M., Houck,J., Hosking,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Matel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nuno,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phoumanong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svitzkas,R., Tector,C., Tyler,D., Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.

Sequencing of *Drosophila melanogaster* genome  
 Unpublished

3 (bases 1 to 302540)  
 Mirra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L., Campbell,K., Hradecky,P., Huang,Y., Kaminker,J.S., Prochuk,S.E., Smith,C.D., Tupy,J.U., Bergman,C.M., Berman,B.P., Carlson,J.W., Celinker,S.E., Clamp,M.E., Drysdale,R.A., Emmert,D., Frise,E., de Grey,A.D.N.J., Harris,N.L., Kronmiller,B., Marshall,B., Milburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E., Shu,S., Smutniak,F., Whitfield,E.J., Ashburner,M., Gelbart,W.M., Rubin,G.M., Mungall,C.J. and Lewis,S.E.

Annotation of *Drosophila melanogaster* genome  
 Unpublished

4 (bases 1 to 302540)  
 Adams,M.D., Celinker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.

Direct Submission  
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

5 (bases 1 to 302540)  
 FlyBase  
 Direct Submission  
 Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA

6 (bases 1 to 302540)  
 FlyBase  
 Direct Submission  
 Submitted (23-JAN-2003) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA

On Sep 13, 2002 this sequence version replaced gi:10726981.  
 Location/Qualifiers  
 1. 302540  
 /organism="Drosophila melanogaster"

gene  
 mRNA  
 CDS  
 gene  
 CDS  
 repeat\_region

/mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /chromosome="X"  
 /note="genome: y[1], cn[1] bw[1] sp[1], Rh6[1]"  
 8252. .9439  
 /locus\_tag="CG1631"  
 /note="Last curated on Wed Jul 03 12:02:51 PDT 2002"  
 /map="19c3-19c3"  
 /db\_xref="FLYBASE:FBgn0031101"  
 /join(8252. .8476,8536. .9439)  
 /locus\_tag="CG1631"  
 /product="CG1631-RA"  
 /db\_xref="FLYBASE:FBgn0031101"  
 /join(8362. .8476,8536. .8972)  
 /locus\_tag="CG1631"  
 /note="CG1631 gene product"  
 /codon\_start=1  
 /product="CG1631-PA"  
 /protein\_id="AAF50920.1"  
 /db\_xref="GI:7295610"  
 /translation="MASPANRDRDQINTADVSDISTSRSELTSDYTSDECTV  
 ADPDTSPQSMKSLASBDSPLBSEQNFCEITSLAHETLCEGDSGSEVSDG  
 EDIDKERYDMFGESESEFGLPTVRLMLSDSDSDNSNEVVGSNPFVETIFLEKDK  
 DQSEKSESKSEKIFQWELFE"  
 complement(47533. .48708)  
 /locus\_tag="CG15462"  
 /note="Last curated on Wed Jul 03 14:50:13 PDT 2002"  
 /map="19c4-19c4"  
 /db\_xref="FLYBASE:FBgn0031102"  
 complement(10147533. .47774,48345. .48708))  
 /locus\_tag="CG15462"  
 /product="CG15462-RA"  
 /db\_xref="FLYBASE:FBgn0031102"  
 complement(10147533. .47774,48345. .48708))  
 /locus\_tag="CG15462"  
 /note="CG15462 gene product"  
 /codon\_start=1  
 /product="CG15462-PA"  
 /protein\_id="AAF50918.1"  
 /db\_xref="GI:7295608"  
 /translation="MNTAVCVYKACAGCEGDSVGVQFTGADSESGREMAATSLSE  
 SKFSASISLRQFVDSRQKQVIEKLGQVAVCRFHSIIDPRDPAITFIDA  
 VMLDFQLACVAVCWEMKLRRLHAAHTVPLPADPTDIDVNVNVNVNVNV  
 NVNGGVGVVDVVDVDEARMDMDMDVLDKAFHCGCQCN"  
 complement(56521. .57280)  
 /locus\_tag="CG12679"  
 /note="Last curated on Wed Jul 03 14:49:36 PDT 2002"  
 /map="19c4-19c5"  
 /db\_xref="FLYBASE:FBgn0031103"  
 complement(56521. .57280)  
 /locus\_tag="CG12679-RA"  
 /product="CG12679-PA"  
 /db\_xref="FLYBASE:FBgn0031103"  
 complement(56578. .57150)  
 /locus\_tag="CG12679"  
 /note="CG12679 gene product"  
 /codon\_start=1  
 /product="CG12679-PA"  
 /protein\_id="AAF50917.1"  
 /db\_xref="GI:7295607"  
 /translation="MNVGALVNLIIILACTYMTSLPPADHPYAFLAASPSVGV  
 GVRSFANPDECGRTFMISATLLEIVLPLANNIFVYSDSGVALVHGSLTFLFY  
 DMIGKISDWDSTETLDMALGNIGTLVIAIKDGNHIFGVAAATLALRYGAIV  
 DSCNGLSNVVELTGNLTGIIALIMVATLEG"  
 58131. .59536  
 /locus\_tag="TBI9654"  
 /map="19c5-19c5"  
 58131. .59536  
 /translation="H-element(1)182"  
 /db\_xref="FLYBASE:FBti10019654"



```

gene      62674..63120
           /locus_tag="TE19655"
           /map="19C5-19C5"
repeat_region 62674..63120
           /transposon="3518(1)183"
           /db_xref="FLYBASE:FB10019655"
gene      complement(71194..71526)
           /locus_tag="CG15460"
           /note="Last curated on Med Jul 03 14:48:49 PDT 2002"
           /map="19C5-19C5"
mRNA      /db_xref="FLYBASE:FBgn0031104"
           complement(71194..71526)
           /locus_tag="CG15460"
           /product="CG15460-PA"
           /db_xref="FLYBASE:FBgn0031104"
           complement(71194..71526)
           /locus_tag="CG15460"
           /note="CG15460 gene product"
           /codon_start=1
           /product="CG15460-PA"
           /protein_id="AA050916.1"
           /db_xref="GI:7295606"
           /db_xref="FLYBASE:FBgn0031104"
           /translation="MSQRPKMPNPNESHSTFLNIPBELQPTRGKPMQKST
           MKGLPSNTVNSAQTAIVNDNLQASVIALMAGMDSILDMEQPNRSSRHERLINE
           LFFSNML"

Query Match      4.9%; Score 39.2; DB 3; Length 302540;
Best Local Similarity 47.5%; Pred. No. 7.6;
Matches 149; Conservative 0; Mismatches 163; Indels 2; Gaps 1;

QY      330 GGTTGATGGCAACCCCAACGCGATCCGAATGGTGGAGTGTCAGTATTCCTTACGCG 389
DB      302342 GGTAAATAATATCATTAACGCGAATGGTGGAGTGAATCTAAATATTAATCTGTTAAAT 302283
QY      390 TTTATTAACCCCTGGAATGTCAACAATAGAAATTTTGTATCGAACCCGAGCGCG 449
DB      302282 ACTTACTTAATCTTTTAAATGCAAGTAACTGAAATTTTAAATGTTAAATTAATTTGCG 302223
QY      450 TAATGGTTGATCTTCACTTACTTACTCAAGAAGAAATGCGCTGTTGATCATGT 509
DB      302222 TAATAGTATATATCTTTTAAACGCGCAAGAAAGAAAGCCATATCTGTTAAATGT 302163
QY      510 CGCGCAATTCACAGATCGCTGGCAAAATTAACCGCTGGGTGAATCAATCCGTGT 569
DB      302162 GAATCAAAACAGCGCATAAATACTGAATACCGCAATTAATCCGTAAACAGATTCCTAT 302103
QY      570 ACTATTTTCACTTA--GAAAAGAGTTTATAGTGGCGGATTTATGATGCTACGCAAA 627
DB      302102 CACATGGCAATGATGATTAATTAATTTGACGCGCAATTTGTGATATCAACGTAGA 302043
QY      628 TTTGGCGTCGAGT 641
DB      302042 TTGCGACTCGGACT 302029

RESULT 12
AC135793/ 128683 bp DNA linear HTG 04-NOV-2003
LOCUS      Oryza sativa (japonica cultivar-group) chromosome 11 clone
DEFINITION OSJNB00017P15, *** SEQUENCING IN PROGRESS ***
AC135793
AC135793.3 GI:38153795
VERSION    HTG; HTGS_PHASE2; HTGS_ACTIVEPIN.
KEYWORDS   Oryza sativa (japonica cultivar-group)
SOURCE      Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
ORGANISM    Eukaryotidae; Magnoliophyta; Liliopsida; Poales; Poaceae;
REFERENCE   1 (bases 1 to 128683)
AUTHORS     Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Ganberger,K., Jones,K.M.,
            Overton II,J., Tselirin,T., Kim,M., Beta,V., Jin,S., Padrosh,D.W.,
            Tallon,L., Xoo,H., Ziemann,V., Hsiao,J., Blunt,S., Vanaken,S.,

```

```

TITLE      Riedmuller,S.B., Uterbach,T., Feldblum,T., Yang,Q., Haas,B.,
            Sub.B., Peterson,J., Quackenbush,J., White,O., Saltsberg,S. and
            Fraser,C.
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 128683)
AUTHORS     Buell,R.
TITLE      Direct Submission
JOURNAL    Submitted (22-OCT-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
            3 (bases 1 to 128683)
REFERENCE   Buell,R.
TITLE      Direct Submission
JOURNAL    Submitted (04-NOV-2003) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            1 128683: contig of 128683 bp in length.
            location/Qualifiers
            1..128683
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="genomic DNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /chromosome="11"
            /clone="OSJNB00017P15"
            /note="japonica cultivar-group"

ORIGIN
Query Match      4.9%; Score 39; DB 2; Length 128683;
Best Local Similarity 11.9%; Pred. No. 7.9;
Matches 48; Conservative 170; Mismatches 185; Indels 0; Gaps 0;

QY      384 TTACGCTTATTAACCGGGAATGCACAAATGGAATTTGGTATGACCCGGA 443
DB      50864 TKATCTTKRKTWTWTCTMTKCCRCGCMWMAWYVYTWYTWTCCTATACMMW 50805
QY      444 GCGGCTAATTTGTTGATCTTCACTTACTCAAGAAGAAATGCGGCTTTGA 503
DB      50804 WMAWACMMWMMWMMWCMYKAYMMRATWTASACTTGGAWYTCYTTGGRKMYRTCT 50745
QY      504 TCATGCGCGCAATTAACGATGCGTGCAGAAAATTAACGCGCTGGTGAATACATC 563
DB      50744 KMYKSTTTGCAWMTKAGKMYGKMTGSCYKRYRAYCYKMSKRRKSYAYR 50685
QY      564 CGTGCTACTATTTTCACTTAAGAAAAGATTTTATAGTGGCGATTTATGATGCTACGC 623
DB      50684 YMYRSAYSWTSYKAYTRRSGNCAAYVYKMAAAWYTWYMKRKGRAAMMYCRM 50625
QY      624 CAAATTTGGCGTGAAGTAAATGATGACCATTAAGCGCGCTTGATCAATCCGGGT 683
DB      50624 WMAAMWRCMTKRTGTGASWKMSPASAAKYMAYCRRKRRKYASTYRKKYKMAWMA 50565
QY      684 GATGTCAGTACCAATTAAGGCGGCGCATCTGTTAAAGGAGGAGCAACGACGAC 743
DB      50564 TWYHGRKMYWMMKMBWTTMMARAKKAYRAAASSMAKRRRTYMSASRRKM 50505
QY      744 CGTTATGCTTGGCCAGTACGATGAGTACACTATTTTCAACC 786
DB      50504 MWMGAMSRRYWMSWMMKRSYMKRYAKCAAYEMWASYY 50462

RESULT 13
AC119922

```



LOCUS AC119922 168422 bp DNA linear HTG 12-MAR-2003  
 DEFINITION Mus musculus clone RP24-252P2, WORKING DRAFT SEQUENCE, 6 unordered  
 pieces.  
 AC119922  
 AC119922 GI:26927735  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 168422)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP24-252P2  
 Unpublished  
 2 (bases 1 to 168422)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,  
 Anderson, S., Barna, N., Bastein, V., Bloom, T., Boguslavsky, L.,  
 Bouguelter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,  
 Charaz, B., Choquet, Y., Colangelo, N., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamet, A., Karatas, A., Kelle, C., Lacroque, K., Lamaras, R.,  
 Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 Maclean, C., MacDonald, P., Major, U., Margis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, U., Menus, L.,  
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunhphang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Roedel, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
 Seaman, J., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 168422)  
 Birren, B., Nusbaum, C., Lander, E., Aboueleil, A., Allen, N.,  
 Anderson, S., Arachchi, H.M., Barna, N., Bastein, V., Bloom, T.,  
 Boguslavsky, L., Bouguelter, B., Camarata, J., Chang, J., Choquet, Y.,  
 Collymore, A., Cooke, P., Cooke, P., Corum, B., Deatellano, K.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, U., Erickson, J., Fato, S.,  
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamet, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,  
 MacDonald, P., Major, J., Manning, U., Matthews, C., McCarthy, M.,  
 Meldrum, U., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunhphang, P., Pierre, N.,  
 Rachupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schupack, R., Seaman, J., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 12, 2003 this sequence version replaced gi:26201661.  
 All repeats were identified using RepeatMasker:  
 Smt, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center Project name: 125309  
 Center Clone name: 252.P.2  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Assembly quality: 16737 bases at least Q40  
 Consensus quality: 167693 bases at least Q30  
 Consensus quality: 167845 bases at least Q20  
 Insert size: 163000; agarose-1p  
 Insert size: 167922; sum-of-coverage  
 Quality coverage: 12.7 in Q20 bases; agarose-1p  
 Quality coverage: 12.3 in Q20 bases; sum-of-coverage  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 11641: contig of 11641 bp in length  
 \* 11642 11741: gap of 100 bp  
 \* 11742 14593: contig of 2852 bp in length  
 \* 14594 14693: gap of 100 bp  
 \* 14694 78832: contig of 64139 bp in length  
 \* 78833 78932: gap of 100 bp  
 \* 78933 115186: contig of 36254 bp in length  
 \* 115187 115286: gap of 100 bp  
 \* 115287 115317: contig of 36031 bp in length  
 \* 115318 151417: gap of 100 bp  
 \* 151418 151418: contig of 17005 bp in length.  
 \* 151418 168422: contig of 17005 bp in length.  
 \*  
 \* Location/Qualifiers  
 \* 1. 168422  
 \* /organism="Mus musculus"  
 \* /mol\_type="genomic DNA"  
 \* /db\_xref="taxon:10090"  
 \* /clone="RP24-252P2"  
 \* /clone\_1lb="RP24-252P2"  
 \* /clone\_1lb="RP24-252P2"  
 \* 1. 11641  
 \* /note="assembly\_fragment"  
 \* clone\_end:8P6  
 \* vector\_side:left  
 \* 11742. 14593  
 \* /note="assembly\_fragment"  
 \* 14694. 78832  
 \* /note="assembly\_fragment"  
 \* 78933. 115186  
 \* /note="assembly\_fragment"  
 \* 115287. 115317  
 \* /note="assembly\_fragment"  
 \* 151418. 168422  
 \* /note="assembly\_fragment"  
 \* clone\_end:17  
 \* vector\_side:right  
 \*  
 \* ORIGIN  
 \* Query Match 4.9%; Score 39; DB 2; Length 168422;  
 \* Best local similarity 46.2%; Pred. No. 8.1;  
 \* Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
 \*  
 \* Db 448 GCTAATGTTGATGATCTTCACTTCACTCAAGAAAGAAATGCGCGCTTGTGATCAT 507  
 \* 138013 GTTGTGTTGTTTCTTCACTATATATAGAAACCAATTTCCCTTTCTTATTTAT 138072  
 \* 508 GTGCGCAATTCAGCATCGTGGCAAAATTCAGCGCTGGAATTCACATCCGCGT 567  
 \* 138073 ATATGCTGAAAAACAGTCTTATCAAGAAATGTTGTCAGGATTAACATCAACAATG 138132  
 \* 568 GTACTATTTTCATTAAGAAAAAGATTTTAGTGGGAGATATATGATGACCAACCA 627

Db 138133 GATGCGAATGGCTTGGGACCAAGGGGCTTGGAACTGGAAAATATCTGGAGTTCTGGGAAT 138192

Qy 628 TTTCGGCTGCAGATTTATATGCGATGACACATTAACCCGCTTGATTAATACGGGGTATC 687

Dd 138193 CTATGTCGACACTGAATCCCATACACATGCGACGACGACAGCTCATTAAGCTTCTG 138252

Qy 688 GTCACTACCAATAAAATGGCCGCATTTGTAAAGCCAA 726

Dd 138253 AGGAGATTAAAGATGATGATGACATCTTTGGCAGAAAAA 138291

RESULT 14	AC107808	187063 bp	DNA	linear	HTG 18-DEC-2003
LOCUS	AC107808				
DEFINITION	Mus musculus chromosome 9 clone RP23-61B14 map 9, *** SEQUENCING IN PROGRESS ***	3 ordered pieces.			

ACCESSION	AC107808
VERSION	AC107808.11
KEYWORDS	GI:39841185
SOURCE	HTG_HTG5_PHASE2; HTG5_FUT10P; HTG5_ACTIVEFIN.
ORGANISM	Mus musculus (house mouse)
	Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 187063)  
Birren, B., Nusbaum, C. and Lander, E.  
Mus musculus chromosome 9, clone RP23-61B14  
Unpublished  
2 (bases 1 to 187063)  
Birren, B., Linton, J., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

REFERENCE  
AUTHORS

2 (bases 1 to 187063)

Birren, B., Linton, L., Nisbun, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Baetien, V., Bogutskiy, L., Boukhalter, B.,  
Brown, A., Camarata, J., Campolongo, A., Chang, J., Chazaro, B.,  
Choepl, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gadya, S.,  
Gilde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Ilev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Laboquie, K., Lamazars, R.,  
Landers, T., Lehoczy, J., Levine, R., Liu, G., McLaren, C.,  
Macdonald, P., Major, T., Marquis, N., Matthews, C., McCarthy, M.,  
Mcwan, P., McKernan, K., Meldrum, U., Meneses, L., Milova, T.,  
Mlenga, V., Murphy, T., Naylor, C., Nguyen, C., Nicoll, R., Norbu, C.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunhthang, P., Platter, N., Pollara, V., Raymond, C.,  
Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppach, R., Seaman, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggilo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembeck, L., Zimmer, A. and Zody, M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (24-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**REFERENCE** 3 (bases 1 to 187063)  
**AUTHORS** Birren,B., Nussbaum,C., Lander,E., Aboueleil,A., Allen,N.,

3 (baes 1 to 187063)

REFERENCES

Bixen, B., Nisbaum, C., Lander, E., Abouellell, A., Allen, N., Anderson, M., Arachchi, H. M., Barn, N., Bastien, V., Bloom, T., Boguslavsky, L., Bourthsgaler, B., Camarot, J., Chang, T., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Cornu, B., Deatrellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, U., Galdyna, S., Graham, L., Grand-Pierre, N., Haez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Macbiter, T., Levine, R., Linblad-Poh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meidrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, C., Phunhang, P., Pierre, N., Ratcupa, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupp, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, B., Vasiliev, H., Venkataraman, V. S., Viet, R., Vo, A., Wilson, B., Wu, X.

**TITLE** Direct Submission  
**JOURNAL** Submitted (18-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Dec 13, 2003 this sequence version replaced gi:38490669.

Wyman, D., Young, G., Zainoun, J., Zemбек, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (18-DEC-2003) Whitehead Institute/MIT Center for Genom  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Dec 13, 2003 this sequence version replaced gi:38490669.  
 All repeats were identified using RepeatMasker:  
 Smit, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 -----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 -----  
 Genome Center  
 -----  
 Center code: MIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 -----  
 Project Information  
 -----  
 Center project name: L20533  
 -----  
 Center clone name: 61\_B\_14

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

*	1	34711:	contig of 34711 bp in length
*	34742	34851:	gap of 100 bp
*	34842	104258:	contig of 69416 bp in length
*	104258	104357:	gap of 100 bp
*	104358	187053:	contig of 82706 bp in length.

```
FEATURES
source
location/Qualifiers
1..187063
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10030"
/chromosome="9"
/map="9"
/clone="RP23-61B14"
/clone_1fb="RPC1-23 Female Mouse BAC"
```

	Query Match	4.98;	Score 39;	DB 2;	Length 187063;
	Best Local Similarity	46.28;	Pred. No. 8.2;	150;	Indels 0;
	Matches 129;	Conservative	0;	Mismatches	0;
Qy	448 GCTAATGGTTTATCTTCATCTTACTTACCAAGAAAGAAATGCCGCTTGCTTTGATCAT				507
Db	57313 GTTTGTTTGGTTTTCCTTAACATAATTAAAGAAACCAATTTCCCTTTCTTTATTAAAT				57372
Qy	508 GTCCGCCAATTCAGCATGCGTGGCAAAAATTACGGCGCTGGCGGTTGAATTCACATCCGNG				567
Db	57373 ATATTGCTGAAAAACAGTCCCTTATCAAGATTGTGTTCAGGGGATTTCAATTCACACAAAG				57433
Qy	568 GTACTATTCTTCATTAGAAAAGAGTTTATAGTGGCGAATATTATTGATGCTACGCCAA				627
Db	57433 GATGCCAATGGCTTGGGACAGGGGCTTGGAACTGGAAATATCTGAGTTCTGGGAAAT				57492
Qy	628 TTGGCGCTCGAAGTTATGCGATGACCAATTAAAGCCGCTTGATCAATACGGGGTGATC				687
Db	57493 CTATGCTGCACACTGTAATCCCATTAACATGCAAGCACAGCCACGCTCATTAAGCTTCG				57552
Qy	688 GTGATGCATTAATAATAGCCGCGATCTTGTAAGGCAAA				726
Db	57553 AGGAGATTAAAGATGATGACACATCTTGGCAGAAAAA				57591

  

RESULT 15	AKI25495/c	AKI25495	2669 bp	MRNA	linear	PRI 09-SEP-2003
LOCUS						
DEFINITION		Homo sapiens cDNA FLJ35306 fis, clone PERIC1000147.				
ACCESSION		AKI25495				
VERSION		AKI25495.1				GT:34531609

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS

1  
Ninomiya,K., Wagatsuna,M., Kanda,K., Kondo,H., Yokoi,T.,  
Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,  
Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,  
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,  
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,  
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,  
Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,  
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,  
Nagahara,K., Masuko,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project

Unpublished  
2 (bases 1 to 2669)

Isogai,T. and Yamamoto,J.

Direct Submission

Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

(E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB) ; cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.

FEATURES  
source location/Qualifiers

1..2669

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="PERIC1000147"

/tissue\_type="pericardium"

/clone\_lib="PERIC1"

/note="Cloning vector: PME18SFL3"

ORIGIN

Query Match 4.9%; Score 38.8; DB 9; Length 2669;

Best Local Similarity 53.2%; Pred No. 6;

Matches 82; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 241 GACCTGTCGATGAGATTTTGATGACGACGACGATTTTGCGCAAAAAGTCCAT 300

DB 268 GGACCTTTCATTCGACATTTTAAAGCTGTGAGATATTTGTTTCATTAGAAAGTCGTG 209

QY 301 ACTTATCCATTTTATCAGGATCCGCTGGTGTGATGCAACCCCAAGCGATCCGAT 360

DB 208 GCTTTTCAAGTAGACATAGTCTCCCGCTTTTGTGTGCTTAATGATGATGACT 149

QY 361 GGTGAGTGTGATTTCCCATTAACGCTTAT 394

DB 148 TTTTCAAGAGCTTTCATTTCTGTACATGTGAT 115

Search completed: April 30, 2004, 15:29:03  
Job time : 3230.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 06:19:01 / Search time 326.63 Seconds  
(without alignments)  
10378.910 Million cell updates/sec

Title: US-10-603-260-2

Perfect score: 798

Sequence: 1 atgttgcacaaagctcta.....ttcaacctgttcagaggt 798

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: N\_Geneseq\_28Jan04:\*

1: Geneseq19808:\*\n2: Geneseq19908:\*\n3: Geneseq20008:\*\n4: Geneseq20018:\*\n5: Geneseq20028:\*\n6: Geneseq20038:\*\n7: Geneseq20048:\*\n8: Geneseq20058:\*\n9: Geneseq20068:\*\n10: Geneseq20078:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	405.8	50.9	717	7	ACA53528 Prokaryot
2	38.2	4.8	6065	8	AAD57245 Human CGD
3	38	4.8	2000	7	ADA71938 Rice gene
4	37.6	4.7	2000	7	ADA71938 Rice gene
5	36.8	4.6	438	5	ABA18862 Human ner
6	36.8	4.6	438	5	ABA18862 Human ner
7	36.8	4.6	438	5	ABA18862 Human ner
8	36	4.5	73465	6	ABQ88161 Human oet
9	35.6	4.5	552	9	ADC90659 E. faeciu
10	35.4	4.4	1119	9	ADB76952 Mouse CLC
11	35.2	4.36	4368	2	AAQ49902 Glutamic
12	35.2	4.4	4512	6	ABI99832 Mouse lsc
13	34	4.3	110000	6	ABO67196-4
14	34	4.3	110000	6	ABO67196-5
15	34	4.3	110000	6	ABO67196-27
16	33.8	4.2	1974	6	AA146803 H influen
17	33.6	4.2	2628	2	AA224773 Human sol
18	33.6	4.2	2628	6	ABK49992 Degenerat
19	33.6	4.2	110000	6	ABA90521_18
20	33.4	4.2	1776	7	ADA70920 Rice gene
21	33.2	4.2	223	7	AA537288 Novel hum
22	33.2	4.2	2622	7	ACA28161 Prokaryot
23	33.2	4.2	82952	6	ABN85766 Arabidops

C 24	33	4.1	395	4	AA185726 Human pol
C 25	33	4.1	3971	7	AA47845 Haemophil
C 26	33	4.1	4942	4	AB105954 Drosophi
C 27	33	4.1	30121	2	AA12063_18
C 28	33	4.1	110000	2	AA12063_17
C 29	32.8	4.1	615	6	ABQ26306 Oligonuc
C 30	32.8	4.1	615	6	ABQ26307 Oligonuc
C 31	32.8	4.1	1109	3	AA455218 Arabidops
C 32	32.8	4.1	1115	3	AA455493 Arabidops
C 33	32.8	4.1	4590	5	AAH24055 Yeast RD
C 34	32.6	4.1	309	4	AAH3611 S. epider
C 35	32.6	4.1	648	7	AB115000 Pathogen
C 36	32.6	4.1	651	4	AAH3615 S. epider
C 37	32.6	4.1	825	6	ABN92979 Staphyloc
C 38	32.6	4.1	3549	4	AAH54658 S. epider
C 39	32.6	4.1	349980	5	AAH66431 Pyrococu
C 40	32.4	4.1	1308	7	ACA36287 Prokaryot
C 41	32.4	4.1	1494	4	AAH52644 E. coli D
C 42	32.4	4.1	1494	7	ACA32678 Prokaryot
C 43	32.4	4.1	1620	2	AAQ03541 Chinese h
C 44	32.4	4.1	1830	7	ACA54254 Prokaryot
C 45	32.4	4.1	2742	5	AAH93830 DNA encod

#### ALIGNMENTS

RESULT 1  
ACA53528 standard; DNA; 717 BP.  
ACA53528:  
19-JUN-2003 (first entry)  
Prokaryotic essential gene #35185.  
Antisense; ds; prokaryotic essential gene; cell proliferation;  
drug design; gene.  
Vibrio cholerae.  
WO200277183-A2.  
03-OCT-2002.  
21-MAR-2002; 2002WC-US009107.  
21-MAR-2001; 2001US-00815242.  
06-SEP-2001; 2001US-00948993.  
25-OCT-2001; 2001US-03428235.  
08-FEB-2002; 2002US-00072851.  
06-MAR-2002; 2002US-0362699P.  
(ELIT-) ELITRA PHARM INC.  
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zykkind JW;  
Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
WPI; 2003-029926/02.  
P-PSDB; ABV99658.  
New antisense nucleic acids, useful for identifying proteins or screening  
for homologous nucleic acids required for cellular proliferation to  
isolate candidate molecules for rational drug discovery programs.  
Claim 14; SEQ ID NO 41398; 1766BP; English.  
The invention relates to an isolated nucleic acid comprising any one of  
the 6213 antisense sequences given in the specification where expression  
of the nucleic acid inhibits proliferation of a cell. Also included are:  
(1) a vector comprising a promoter operably linked to the nucleic acid  
encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a gene that influences the activity of  
 CC the gene product or that has an activity against a biological pathway; (8)  
 CC required for proliferation, or that inhibits cellular proliferation; (9)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences

Sequence 717 BP, 175 A, 165 C, 187 G, 190 T, 0 U, 0 Other;

Query Match 50.9%; Score 405.8; DB 7; Length 717;

Best Local Similarity 73.1%; Pred. No. 4e-119; Matches 521; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

85 ATGATGTGCTATCGATATGATTTGTGCTGCTGCGCCGAAATCTATCCAGGTTTAA 144  
 1 ATGATGTGCTATCGATATGATTTGTGCTGCTGCGCCGAAATCTATCCAGGTTTAA 60  
 145 CTGGTGAACCGCTTATTCGAAATGCGCAATTTGTGTAATGCGCATTTGCGCGG 204  
 61 TTGGTCAACGCGCAATTCGAAATGCGCAATTTGTGTAATGCGCATTTGCGCGG 120  
 205 ATAGTGTATGAGCAATGATGACCGCTCATGATGAGCAATGATGAGGATTTTAT 264  
 121 TGGGTGTATGAGGATCTCAGCGCAATGATGAGGATTTTAT 180  
 265 GCAAGCAGACGATTTGCGGCAAAAGTCATTTATCTTATTTATGACGAT 324  
 181 TCCGCGCGTGGGCGATCTGCGCGCAAAATCTTATCCCAATCTTATGATGAT 240  
 325 CCGCTGGTGTATGAGCAATTCGCAATGCGCAATTTGTAATGCGCATTTTCCAT 384  
 241 CTTTGTGATGAGGATCTTAAACGCAATTCGCAATTTGTAATGCGCATTTTCCAT 300  
 385 TACGCTTATTAACCGTGAATGCAAAATGAGATTTTGTATGACCGCGAG 444  
 301 TATGCGCTGTGATATTCGCAATGCAATGAGATTTGTAATGAGGATTTGAGG 360  
 445 CCGCTTATTTGTTGATCTTATCTTATCTTATCTTATCTTATCTTATCTTAT 504  
 361 CGTGTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 420  
 505 CATGTGCGCAATTTGAGATGCGTGGCAAAATTAATGCGCGTGGTGAATGACATCC 564  
 421 CATGTGCGCAATTTGAGATGCGTGGCAAAATTAATGCGCGTGGTGAATGACATCC 480  
 565 GTGGTATTTTCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 624  
 481 GTGGTATTTTCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 540  
 625 AATTTTGGGTGCAATTTATGCAATGCAATTTATGCAATTTATGCAATTTATGCA 684  
 541 AATTTTGGGTGCAATTTATGCAATGCAATTTATGCAATTTATGCAATTTATGCA 600

Qy 685 ATGCTCAGTACCAATAAATGCGCATCTGTGAAGCAAGCAACCAACGACACC 744  
 Db 601 ATGCTCAGTACCAATAAATGCGCATCTGTGAAGCAAGCAACCAACGACACC 660  
 Qy 745 GTTTATGCTTTTCCAGTCAATGATGATGATTTTCAAACTGTTAGAGG 797  
 Db 661 GTTTATGCTTTTCCAGTCAATGATGATGATTTTCAAACTGTTAGAGG 713

RESULT 2  
 AAD57245/c  
 ID AAD57245 standard, cDNA, 6065 BP.  
 XX  
 AC AAD57245;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Human CGDD-25 cDNA.  
 XX  
 KW Human; cell growth, differentiation and death protein; CGDD; leukaemia;  
 KW neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;  
 KW muscular disorder; myotonic dystrophy; catatonias; endocrine disorder;  
 KW diabetes; Grave's disease; cancer; immunological disorder; scleroderma;  
 KW systemic lupus erythematosus; allergy; Crohn's disease; renal disorder;  
 KW gastrointestinal disorder; Goodpasture's syndrome; infection; cirrhosis;  
 KW cardiovascular disorder; atherosclerosis; hepatic disease; transgenic;  
 KW transgenic animal; gene therapy; neuroprotective; relaxant; cytostatic;  
 KW dermatological; immunosuppressive; cerebroprotective; anticonvulsant;  
 KW antibacterial; antiparasitic; fungicide; virucide; uropathic; cardiac;  
 KW protozoacide; nootropic; gene; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1442..4129  
 FT /tag= a  
 FT /product= "Human CGDD-25 protein"  
 EN  
 XX MO203050253-A2.  
 PD 19-JUN-2003.  
 XX  
 PF 04-DEC-2002; 2002WC-US039133.  
 XX  
 PR 07-DEC-2001; 2001US-0340747P.  
 PR 20-DEC-2001; 2001US-0342761P.  
 PR 15-JAN-2002; 2002US-0349705P.  
 PR 06-FEB-2002; 2002US-0354764P.  
 PR 12-FEB-2002; 2002US-0356216P.  
 XX  
 RA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Griffin JA, Ramkumar J, Emerling BM, Kable AE, Elliott VS,  
 PI Margus JP, Baughn MR, Gorvay AE, Yue H, Lee EA, Becha SD, Tang YT,  
 PI Tran UK, Swarnakar A, Lee S, Ison CH, Hatalla AJA, Tran B,  
 PI Sprague WM, Lee SY, Khare R, Gandhi AR, Gietzen KJ, Bhatia U,  
 PI Burrill JD, Blake UT, Ho A, Zheng W,  
 XX  
 DR WPI; 2003-532903/50.  
 DR P-PSDB; AAE37936.  
 XX  
 PT New CGDD polypeptides, useful for diagnosing, preventing, and treating  
 PT disorders associated with an abnormal expression or activity of CGDD,  
 PT e.g. neuromuscular, immunological, cardiovascular disorders, cancer  
 PT and/or infections.  
 XX  
 PS Claim 5; Page 289-290; 299pp; English.  
 CC The present invention relates to novel cell growth, differentiation and  
 CC death (CGDD) proteins and polynucleotides encoding them. The sequences of  
 CC the invention are useful in diagnosing, preventing and treating disorders  
 CC associated with an abnormal expression or activity of CGDD such as  
 CC neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's

CC disease), muscular disorders (e.g. myotonic dystrophy, catatonias),  
 CC endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g.  
 CC leukemia, cervical or breast cancers), immunological disorders (e.g.  
 CC scleroderma, systemic lupus erythematosus, allergies), gastrointestinal  
 CC disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's  
 CC syndrome), infections (e.g. viral, bacterial, fungal, parasitic,  
 CC protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis)  
 CC and hepatic diseases (e.g. cirrhosis). The polynucleotides can be used to  
 CC create humanised animals or transgenic animals to model human diseases.  
 CC The invention is also used in gene therapy. The present sequence is human  
 CC CGDD-25 cDNA

Sequence 6065 BP, 2249 A, 1216 C, 1149 G, 1451 T, 0 U, 0 Other;

Query Match 4.8%; Score 38.2; DB 8; Length 6065;

Best Local Similarity 52.9%; Pred. No. 0.62; Matches 82; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 240 AGAAGCTGTGATGAGCATTTGATGACGACGACGATTTGTCGCAAAAGTCCA 299  
 DB 290 AGGACCTTTCATTCGCTATTTAAAGCTGTGAGATTTGTTTCAATGAAAGTCT 231  
 QY 300 TACTTATCTAATTTATGACGATCCGCTGTGTATGAGCAACCCCAACGATCCGAA 359  
 DB 230 GCGTTTCAAGCAATCAATGCTCCGCTTTTGTGTGCTTAAATGATTGTAC 171  
 QY 360 TGGTTGAGTGTGATGATTTCCCATTCAGCTTTAT 394  
 DB 170 TTTTTCAGAGACTTCATTTCTGTACATGGAT 136

# RESULT 3

ADA71938/c  
 ID ADA71938 standard; DNA; 2000 BP.

AC ADA71938;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

XX Plant; bacterial infection; fungal infection; viral infection; rice;  
 XX gene; de.

XX Oryza sativa.

XX WO2003000838-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-1B001105.

XX 22-JUN-2001; 2001WO-1B001105.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 XX Katagiri F, Qian S, Tao Y, Whitcham S, Xie Z, Zhu T, Zou G;  
 XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to  
 XX pathogenic infection for conferring resistance or tolerance to a plant to  
 XX bacterial, fungal or viral infection by determining or detecting plant  
 XX gene expression.

XX Claim 27; SEQ ID NO 5263; 8999p; English.

XX The present invention relates to a method (M1) for identifying genes  
 XX involved in plant resistance or response to pathogenic infection. M1  
 XX comprises identifying a gene whose expression is significantly altered in  
 XX the incompatible interaction of plant gene expression relative to  
 XX expression of the gene in an uninfected plant, in a mutant plant that

CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.

Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.8%; Score 38; DB 7; Length 2000;

Best Local Similarity 9.6%; Pred. No. 0.42; Matches 74; Conservative 335; Mismatches 355; Indels 4; Gaps 2;

QY 19 TATTACATTTGACCATCATCATGACCTTACCTTAATAAGCCGTTGTTATTAGGAA 78  
 DB 855 TMMWRYMTTYCYAATCAKCKYKMTMTTACMRATSRMRPAMAGRWRYRMYK 796  
 QY 79 GCGATTATGATGTGACATCATGATATGATTTGCTGCGCTTCCGCGGAAT-CTATCCA 137  
 DB 795 AYMMRWRCWKAGARWKRMYRMYKRYMYRYMYKMTTMMWRYMRMSYMMWGMGR 736  
 QY 138 GGTTTTACTGTGAAGCTCTATCCAAATCCGAGATTTGTATATGCGCATTC 197  
 DB 735 MRMWRYMRWRCWKATKYSASRMTBRASRYRMRWRYRMYRMYRMYRMYRMYR 676  
 QY 198 TGGCGGATATGTATGACGACGATGACGCTCATGTGTGAGAACCTGTGATGAGA 257  
 DB 675 ARMSKRWKAGASMSKSMYRWRGASWYSKYSKCKCTTMYMTSYSTGYMYSS 616  
 QY 258 TTTTGAAGACGACGACGACGATTTTGTGGGAAAAGTCATATCTATCTAATTTAT 317  
 DB 615 YKMSWTSKMSYGNKMTCTMYTSMKSTRSKGRSNGSRMYRMYRMYRMYRMYR 556  
 QY 318 CAGGATCCGCTGTTATGACCAACC--CAACGCGATCCGATGTTGAGTGCAT 374  
 DB 555 WKTWRRCMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 496  
 QY 375 TATTTCCATTAAGCTTATTAACCCGCGAATGTCACAAATAGAAATTTGGTAT 434  
 DB 495 CSTWRRGYCKACKCCYACWFAYSGMMYRYSKMRSTYKMSMTYKCKCSMK 436  
 QY 435 CGAAGCCGAGCGGCTTATGTTGATCTTCACTTACTTACTCAAGAAAGAAATGCGCT 494  
 DB 435 YGAKGCGYCGMMYRYSYGNKMTCTMYTSMKSTRSKGRSNGSRMYRMYRMYRMYR 376  
 QY 495 GCGTTTGAATGATGCGGCAATTCAGCATGCGTGGCAAAATTCAGCGCTGCGTTGA 554  
 DB 375 WYTYAKYMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 316  
 QY 555 ATACACATCCGCTGTTATTTTCAATTAAGAAAGATTTTGTAGGCGGATTTATGGA 614  
 DB 315 CWRYATCYGCCYRKGWYSRSMRTAGMKRMSRMSRMSRMSRMSRMSRMSRMSR 256  
 QY 615 TGCTTACGCAAAATTTGGCGTGAAGTATGATGACCATTAACGCGCTTGATCA 674  
 DB 255 YGARSSTGWSRAAKRTYGYSTSRRAKMRACRYASCRYSRYSYCGCGSSSKX 196  
 QY 675 TACCGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 734  
 DB 195 KYMSKSGSMWTSNCSNCSNCSNCSNCSNCSNCSNCSNCSNCSNCSNCSNCSNCSN 136  
 QY 735 ACCAGCACCGTTTATGCTTGTGCGAGTATGATGATGATGATGATGATGATGATGAT 782  
 DB 135 KYSCGCTKYSGYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYR 88

# RESULT 4

ADA71938  
 ID ADA71938 standard; DNA; 2000 BP.

AC ADA71938;

XX 20-NOV-2003 (first entry)





[illegible]

PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250391P.
PR	01-DEC-2000;	2000US-0251160P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251865P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI, 2001-541565/60.	
DR	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system cancers	
PR	and metastases.	
PS	Disclosure; SEQ ID NO 11193; 1701bp + Sequence Listing; English.	
XX		
CC	The invention relates to novel genes (ABAI1004-ABA21534) and proteins	
CC	(ABAI4678-ABAI8001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful in	
CC	the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and	
CC	ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,	
CC	breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune	
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic	
CC	anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,	
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)	
CC	cardiovascular disorders such as myocardial ischaemias; (d) wound healing	
CC	; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)	
CC	infectious diseases such as viral, bacterial, fungal and parasitic	
CC	infections. Note: The sequence data for this patent did not form part of	
CC	the printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences	
SQ	Sequence 438 BP; 130 A; 78 C; 104 G; 126 T; 0 U; 0 Other;	
Query March	4.6%; Score 36.8; DB 5; Length 438;	
Best Local Similarity	59.6%; Pred. No. 0.48; Indels 0; Gaps 0;	
Matches	62; Conservative 0; Mismatches 42;	
Qy	357 GAATGTTGGAGTGTCTCAGTATTTCCTTACCGCTTATTAAACCGTGGAATGCATAACA	416
Yb	204 GGAAAGATAGAGTCTCAGTATTTCAGATTACATTACGCTTATTTCGAATGGAATGATAAACA	263
Qy	417 AATTGAAGATTTTGGTATTCGACCCGAGCGCGCTAATTGGTTG	460
Db	264 AACCTAACATTTTGGTAGTCTCTGGGTTGGCTCTTTTAACTTTG	307
RESULT 6		
ID	ABAI1863 standard; DNA; 438 BP.	
XX	ABAI1863;	
AC	ABAI1863;	
DT	23-JAN-2002 (first entry)	
XX		
DE	Human nervous system related polymnucleotide SEQ ID NO 11194.	
KM	Human; noctropic; neuroprotective; cytosstatic; dermatological; virucide;	
KM	immunopressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;	
KM	immunosuppressive; antiscintigraphic; antianemic; antidiabetic; cancer;	

KM antineumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KM antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KM neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 OS Homo sapiens.  
 XX WO200159063-A2.  
 XX PD 16-AUG-2001.  
 XX PF 17-JAN-2001; 2001WO-US001334.  
 XX PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205151P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-022513P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225265P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225477P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226686P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-023497P.  
 PR 25-SEP-2000; 2000US-023498P.  
 PR 25-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 20-OCT-2000; 2000US-0242212P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249247P.  
 PR 17-NOV-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0251989P.  
 PR 05-DEC-2000; 2000US-0251990P.  
 PR 05-DEC-2000; 2000US-0251991P.  
 PR 08-DEC-2000; 2000US-0251866P.  
 PR 08-DEC-2000; 2000US-0251867P.  
 PR 08-DEC-2000; 2000US-0251869P.

PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
XX Disclosure; SEQ ID NO 11194; 1701bp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at [http://wipo.int/pdb/published\\_pdb\\_sequences](http://wipo.int/pdb/published_pdb_sequences)  
XX  
SQ Sequence 438 BP; 130 A; 78 C; 104 G; 126 T; 0 U; 0 Other;  
  
Query Match 4.6%; Score 36.8; DB 5; Length 438;  
Best Local Similarity 59.6%; Pred. No. 0.48;  
Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
  
QY 357 GATGGTTCAGTGTGATGATTTCCATTCAGCTTATTAACCGGTGAATGTCAACA 416  
Db 204 GGAAGATATGAGTGGTGAATTCATTAACCTTATTTCACTTGAATGATAAACA 263  
QY 417 AATGAAGATTTTGGTATGACCCGAGCGGCTTAATGGTTG 460  
Db 264 AACCTAACATTTGGTAGTCTCGGTTGCTCTTTACTTTG 307  
  
RESULT 7  
ABA18866  
ID ABA18866 standard; DNA; 439 BP.  
XX  
XX ABA18866;  
AC  
XX  
XX 23-JAN-2002 (first entry)  
DT  
XX  
XX Human nervous system related polynucleotide SEQ ID NO 11197.  
DE  
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
XX immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;  
XX antiparkinsonian; antischistosomal; antianaemic; antithrombotic; cancer;  
XX antitubercular; hepatotropic; cerebroprotective; anti-inflammatory;  
XX antiallergic; antidiabetic; antitumor; anticomvulsant; antifungal;  
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200159063-A2.  
PN  
XX  
XX 16-AUG-2001.  
PD  
XX

PF 17-JAN-2001; 2001WO-US001334.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218280P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234232P.  
PR 21-SEP-2000; 2000US-0234233P.  
PR 25-SEP-2000; 2000US-0234574P.  
PR 25-SEP-2000; 2000US-0234575P.  
PR 26-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.

[illegible]

```

PT useful for preventing, diagnosing and/or treating nervous system cancers
PI and metastases.
PS Disclosure; SEQ ID NO 11197; 1701bp + Sequence Listing; English.
XX The invention relates to novel genes (ABR11004-ABR21534) and proteins
CC (ABR11678-ABR18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences
CX
SQ Sequence 439 BP; 130 A; 79 C; 104 G; 126 T; 0 U; 0 Other;
Query Match 4.6%; Score 36.8; DB 5; Length 439;
Best Local Similarity 59.6%; Pred. No. 0.48; Mismatches 0; Gaps 0;
Matches 62; Conservative 0; Indels 42;
OY 357 GAATGTTGGAGTGTCACTATTTCCTTAACGCTTATTAACCGTGGAATGCATAACA 416
DB 205 GGAAAGTAGAGTCTGAGTATTCACACTTAACGCTTATTTCAACTTGGAATGATMAACA 264
OY 417 AATAGAAGATTTTGGTATCGACC CGCGCGGCTAATGGTTTG 460
DB 265 AACCTAACATTTTGTGTAGTCTGTGGGTTTGCTCTTTAACTTTG 308
RESULT 8
ABR08161/c
ID ABR08161 standard; cDNA; 73465 BP.
XX
AC ABR08161;
XX
DT 18-SEP-2002 (first entry)
XX
DE Human osteoblast differentiation related cDNA SEQ ID NO 68.
XX
KW Human; osteoblast; stem cell differentiation; bone tissue deposition;
KM osteoporosis; osteopathic; ss.
XX
OS Homo sapiens.
XX
PN WO200250301-A2.
XX
PD 27-JUN-2002.
XX
PF 18-DEC-2001; 2001WO-US048276.
XX
PR 18-DEC-2000; 2000US-0255882P.
XX
PR 24-APR-2001; 2001US-0285691P.
XX
PA (GENE-) GENE LOGIC INC.
PA (PROC ) PROCTER & GAMBLE CO.
XX
PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A,
PI Mertiz L;
XX
DR WPI; 2002-557663/59.
XX
PT Use of genes and their expression profiles associated with osteoblast
PT differentiation for screening modulators bone formation; for diagnosing
or treating e.g. osteoporosis, or as markers for the differentiation

```

PT process.  
XX  
PS Claim 1; SEQ ID NO 68; 789p + Sequence Listing; English.  
XX  
CC The invention relates to genes and their expression profiles are used  
CC for: (a) screening modulators of precursor stem cell differentiation into  
CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal  
CC deposition of bone tissue, abnormal rate of osteoblast formation or  
CC osteoporosis; or (c) treating or monitoring treatment of the conditions  
CC cited in (b), or monitoring the progression of bone tissue deposition.  
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid  
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-  
CC induced abnormalities in bone formation or bone loss, conditions that  
CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),  
CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome  
CC or fibrous dysplasia. The present sequence is that of an osteoblast  
CC differentiation associated cDNA marker of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [http://wipo.int/publ/published\\_pct\\_sequences](http://wipo.int/publ/published_pct_sequences)  
XX  
SQ Sequence 73465 BP; 21307 A; 15130 C; 15840 G; 21188 T; 0 U; 0 Other;  
XX  
Query Match 4.5%; Score 36; DB 6; Length 73465;  
Best Local Similarity 51.2%; Pred. No. 11;  
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
XX  
QY 20 ATTTCATTGACATCATCATGCACTTACCTTAATAAGCCCTGTTTATTAGGGAAG 79  
Db 72386 ATACATATTTTACATTAATTAATGACAAACGACATTAATGAGTTCTTGTGAT 72327  
QY 80 CCATTATGATTTGTCTACTATCATGATGATTTGTCTGCGCTTGCGCGAATCTATCCAGG 139  
Db 72326 GCAATTAAGAGAGTGTCTTCTCATAGATTAATTAATTTGTATTTCTTATTTCTTACTGC 72267  
QY 140 TTTTACTGTTGAACGCTCTATCCAAATCGGCGAGATTGTGGT 183  
Db 72266 TTGCTGCGCTGTAAACGTTTATAAATGAGTTCTTTGGGCT 72223  
XX  
RESULT 9  
AD90659  
ID ADB90659 strand; DNA; 552 BP.  
XX  
AC ADB90659;  
XX  
DT 01-JAN-2004 (first entry)  
XX  
DE E. faecium DNA sequence SEQ ID 286.  
XX  
KW ds; gene; urinary tract infection; bacteraemia; endocarditis; wound;  
XX abdominal-pelvic infection.  
XX  
OS Enterococcus faecium.  
XX  
PN US6583275-B1.  
XX  
PD 24-JUN-2003.  
XX  
PF 30-JUN-1998; 98US-00107532.  
XX  
PR 02-JUL-1997; 97US-0051571P.  
PR 14-MAY-1998; 98US-0085598P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI: 2003-799836/75.  
XX P-PsDB; ADB94313.  
XX  
PT New isolated nucleic acid derived from Enterococcus faecium encoding an  
PT Enterococcus faecium polypeptide useful for detection, prevention and

PT treatment of a pathological condition resulting from a bacterial  
PT infection.  
XX  
PS Example 1; SEQ ID NO 286; 243bp; English.  
XX  
CC The invention relates to an isolated nucleic acid derived from  
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
CC one of 10 fully defined sequences given in the (or comprising 40  
CC sequential nucleotides chosen from any of the nucleic acids, its  
CC complement or sequences hybridizing to it). Also included are a  
CC recombinant vector comprising the nucleic acid operably linked to  
CC transcription regulatory element, a cell comprising the vector and a  
CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
CC The nucleic acids is useful for diagnosing pathological conditions  
CC resulting from E. faecium bacterial infection (e.g. urinary tract  
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
CC infection) and for screening drugs such as agonists and antagonists. The  
CC nucleic acid is useful for recombinant production of Candida albicans -  
CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
CC and vaccines containing the nucleic acid are useful for preventing or  
CC treating Enterococcus faecium infections. The present sequence represents  
CC one if the disclosed E. faecium nucleic acids.  
XX  
SQ Sequence 552 BP; 174 A; 96 C; 135 G; 147 T; 0 U; 0 Other;  
XX  
Query Match 4.5%; Score 35.6; DB 9; Length 552;  
Best Local Similarity 51.2%; Pred. No. 1.3;  
Matches 83; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
XX  
QY 237 TGGAGAACCTGTGATGAGGATTTTATGACAGGAGACGATTTTGTGCGAATAAGT 296  
Db 294 TAGTCAACATCTCAGAGATATCTATGTTTGGGTGGAGCGGATTTTTCMAAGACT 353  
QY 297 CCATTCTTATCTTAATTTTATGACGATCGGCTGTTATGCGAACCCCAAGCGCATCC 356  
Db 354 TCTTCTTACACCGGTGCAATCTGGCGGACATGATGATGACATGTTTTCMAAGACT 413  
QY 357 GAATGTTGGAGTGTGATTTTCCATTACGCTTTATTAA 398  
Db 414 GTATATGGGGATATCATTTTTCAGACCTTGGCTTATAGA 455  
XX  
RESULT 10  
ADB76952  
ID ADB76952 strand; DNA; 1119 BP.  
XX  
AC ADB76952;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Mouse CLCA4 gene SEQ ID NO:20.  
XX  
KW respiratory; gastrointestinal; nephrotropic; respiratory disease;  
XX nephritis; digestive disease; mouse; CLCA4; ds; gene.  
XX  
OS Mus sp.  
XX  
PN Key Location/Qualifiers  
XX CDS 1..1119  
XX FT /tag= a  
XX FT /product= "CLCA4"  
XX FT /note= "No start/stop codon given"  
XX  
FN WO2003062426-A1.  
XX  
PD 31-JUL-2003.  
XX  
PF 20-JAN-2003; 2003WO-JP000408.  
XX  
PR 21-JAN-2002; 2002JP-00012180.  
PR 21-FEB-2002; 2002JP-0004981.  
PR 07-MAR-2002; 2002JP-00061668.

```

XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX Nakanishi A, Morita S;
XX
XX WPI; 2003-598752/56.
XX
XX DR P-PSDB; ADB76951.
XX
XX PT Proteins for treatment, diagnosis and prevention of respiratory disease,
XX nephritis, digestive diseases.
XX
XX PS Claim 12; Page 104; 113pp; Japanese.
XX
XX CC The invention relates to a novel protein comprising a sequence selected
XX from two protein sequences each 924 amino acids in length. A protein of
XX the invention has respiratory, gastrointestinal, and nephrotropic
XX activity. The protein is useful for treatment, diagnosis and prevention
XX of respiratory disease, nephritis, and digestive diseases. The present
XX sequence encodes mouse ClCA4.
XX
XX SQ Sequence 1119 BP; 352 A; 245 C; 261 G; 261 T; 0 U; 0 Other;
XX
XX Query Match 4.4%; Score 35.4; DB 9; Length 1119;
XX Best Local Similarity 50.3%; Pred. No. 2.1;
XX Matches 87; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
XX
XX QY 601 GCGGATTTTATGATGCTTACGCCAATTGGCGTGAAGTTATGCGATGACATTA 660
XX DB 898 GCTGATGCTACTAGAGAGTCCCTTCAGCTTGAAGAGAGGTCCTGTGACCAACGT 957
XX QY 661 CGCGGCTTGCATCAATACCGGGGTGATGTCAGTACCAATTAATGCGGCATCTGTAA 720
XX DB 958 CGCTGGCTGACACACACTGTCGTTATTGACAGCACCGGTGAAAGACACATCTTCT 1017
XX QY 721 GCGAAGAGAGCCAAACCGCCGCTTATGCTTCCAGTCATGAAGTCAC 773
XX DB 1018 GTACACCTGAGCAACACAGCCCTGCCATTCACTCAGGAGTCCAAAGAAAC 1070
XX
XX RESULT 11
XX AAQ49902/C
XX ID AAQ49902 standard; cDNA; 4368 BP.
XX
XX AC AAQ49902;
XX
XX XX
XX DT 14-MAY-2003 (revised)
XX DT 04-MAY-1994 (first entry)
XX
XX DE Glutamic acid receptor.
XX
XX XX
XX KW Glutamic acid receptor; nerve; synapse; synapse plasticity; cerebellum;
XX cell necrosis; ischemia; ss.
XX
XX OS Mus musculus.
XX
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 1..4368
XX FT /*tag= a
XX FT /product= "glutamic_acid_receptor"
XX
XX FN JP05239098-A.
XX
XX XX
XX PD 17-SEP-1993.
XX
XX XX
XX PF 26-FEB-1992; 92JP-00039563.
XX
XX PR 26-FEB-1992; 92JP-00039563.
XX
XX PA (MITU ) MITSUBISHI KASEI CORP.
XX
XX WPI; 1993-331429/42.
XX
XX DR P-PSDB; AAR42054.
XX
XX

```

```

PT New protein for elucidation of nerve information transfer by synapse -
PT comprises glutamic acid receptor of specified aminoacid sequence.
XX
XX PS Claim 1; Page 11; 15pp; Japanese.
XX
XX CC The sequence was transformed into E. coli several clones coding the new
XX glutamic acid receptor were obtained. The sequence is useful for the
XX elucidation of nerve information transfer by synapses, expression of
XX synapse plasticity and nerve cell necrosis caused by cerebral ischemia
XX and also for the development of new drugs. In the specification the
XX length of the sequence is given as 4392 bp The actual number of bases is
XX 4368. The sequence may have a line missing or has been numbered
XX incorrectly. (Updated on 14-MAY-2003 to correct PS field.)
XX
XX SQ Sequence 4368 BP; 1151 A; 1127 C; 1090 G; 1000 T; 0 U; 0 Other;
XX
XX Query Match 4.4%; Score 35.2; DB 2; Length 4368;
XX Best Local Similarity 49.0%; Pred. No. 4.8;
XX Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
XX
XX QY 234 TGTGAGAGACCTGTGATGAGATTGTTGATGACGACGACGATTTGTGGCAAAA 293
XX DB 553 TGATGAAGTCTCTGTACCGAGGAAGATGGTGTGACCAAGAGAGACATGCCAGTCAT 494
XX QY 294 AGTCCATACCTTACCTAATTTTATCAGCGAATCCGCTGTGATGGAACCCCAACCGCA 353
XX DB 493 AGTCTGATGATCTTTCAGCATGACCGTGTCTGTCTGTGATGGAAGCTCCAACTGGA 434
XX QY 354 TCCGATGGTGTGAGTGTGATGATTTCCATTAAGCTTTAATTAACCCGTGAATGTCAA 413
XX DB 433 AGAAGCGAGATGTGATGATCCCTTGACGCCATGATCATGATGACCCCATGATGCCCA 374
XX QY 414 ACAATAGAGA 425
XX DB 373 AGATGGGATGA 362
XX
XX RESULT 12
XX AB199832/C
XX ID AB199832 standard; cDNA; 4512 BP.
XX
XX AC AB199832;
XX
XX XX
XX DT 07-MAR-2002 (first entry)
XX
XX DE Mouse ischemic condition related cDNA sequence SEQ ID NO:952.
XX
XX XX
XX KW Mouse; ischemia; compressive ischemia; occlusive ischemia;
XX vasospastic ischemia; ischemic condition; ischemic disease; ss.
XX
XX OS Mus musculus.
XX
XX XX
XX FN WO20018188-A2.
XX
XX PD 22-NOV-2001.
XX
XX XX
XX PF 18-MAY-2001; 2001WO-JP004192.
XX
XX PR 18-MAY-2000; 2000JP-00145977.
XX
XX PA (UNIV-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
XX XX
XX FI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
XX DR WPI; 2002-034733/04.
XX
XX DR P-PSDB; ABB57341.
XX
XX PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
XX expression levels of particular genes defined in the specification or by
XX determining the expression profile of a gene group comprising these
XX genes.
XX
XX Claim 2; Page 2406-2415; 2690pp; English.
XX

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:27 ; Search time 2265.3 Seconds  
(without alignments)  
10519.601 Million cell updates/sec

Title: US-10-603-260-2

Perfect score: 798  
Sequence: 1 atgttgcacacaaagtctta.....ttcaacctgttaccagagc 798

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estcov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_pig:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41.6	5.2	1048	9	AL545951 AL545951
2	41	5.1	486	14	CF088325 QHM18L03
3	40.8	5.1	746	29	CC876255 ZMMBBD019
4	40.8	5.1	826	28	BZ723869 PUBNSO9TD

Result No.	Score	Query Match	Length	DB ID	Description
5	40.8	5.1	919	28	CC389867 CC389867
6	40.4	5.1	438	14	CF090249 CF090249
7	40.4	5.1	464	14	CF087071 CF087071
8	40.2	5.0	638	28	AZ396257 AZ396257
9	39.8	5.0	401	9	AJ539921 AJ539921
10	39.8	5.0	435	14	CF090254 CF090254
11	39.6	5.0	437	13	BX500256 BX500256
12	39.4	4.9	447	14	CF091479 CF091479
13	39.4	4.9	454	14	CF0908121 CF0908121
14	39	4.9	401	14	CF096847 CF096847
15	39	4.9	401	9	AJ540178 AJ540178
16	38.8	4.9	418	14	CF090542 CF090542
17	38.8	4.9	427	14	CF091120 CF091120
18	38.8	4.9	437	14	CF090736 CF090736
19	38.8	4.9	440	14	CF088879 CF088879
20	38.8	4.9	452	14	CF087459 CF087459
21	38.8	4.9	454	14	CF090001 CF090001
22	38.8	4.9	468	14	CF091500 CF091500
23	38.8	4.9	472	14	CF088587 CF088587
24	38.8	4.9	487	14	CF087094 CF087094
25	38.8	4.9	520	14	CF090018 CF090018
26	38.6	4.8	274	14	CF096559 CF096559
27	38.2	4.8	417	14	CF091407 CF091407
28	38.2	4.8	429	14	CF088180 CF088180
29	38.2	4.8	433	14	CF087501 CF087501
30	38.2	4.8	432	14	CF087987 CF087987
31	38.2	4.8	542	14	CF089042 CF089042
32	38.2	4.8	581	14	CF091617 CF091617
33	38.2	4.8	711	10	BF693394 BF693394
34	38.2	4.8	9744	29	AY405009 AY405009
35	38	4.8	635	14	CF094197 CF094197
36	37.8	4.7	374	14	CF087834 CF087834
37	37.8	4.7	409	14	CF090830 CF090830
38	37.8	4.7	486	14	CF090109 CF090109
39	37.4	4.7	283	14	CF095069 CF095069
40	37.2	4.7	434	14	CF089989 CF089989
41	37.2	4.7	448	14	CF087103 CF087103
42	37.2	4.7	454	14	CF088135 CF088135
43	37.2	4.7	460	13	BX485370 BX485370
44	36.8	4.6	440	10	BE081047 BE081047
45	36.6	4.6	331	14	CF088421 CF088421

## ALIGNMENTS

RESULT 1  
LOCUS AL545951 1048 bp mRNA linear EST 31-MAY-2003  
DEFINITION AL545951 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
Clone CSOD10237B12 5-PRIME, mRNA sequence.  
ACCESSION AL545951  
VERSION AL545951.2 GI:31267786  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Bhakayote, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 1048)  
JOURNAL Full-length cDNA libraries and normalization  
COMMENT On Feb 15, 2001 this sequence version replaced GI:12876614.  
Contact: Genoscope - Centre National de Sequencage  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Parade Avenue Genoscope sequence ID: CSOD10237B06P1.

## FEATURES

source 1..1048

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS01023YB1.2"  
/issue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-collg (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match 5.2%; Score 41.6; DB 9; Length 1048;

Best Local Similarity 47.3%; Pred. No. 0.5; Indels 0; Gaps 0;

Matches 80; Conservative 12; Mismatches 77; Indels 0; Gaps 0;

Db 341 ACCCCAAACGCGATCCGATGTTGAGTGCAGTATTCCTTACGCTTATTAACC 400  
536 AACCTCAAGAGAACCAATTAAGCTTGAGTGCCTTAATTTTAAACGATTCATTAAC 595  
Qy 401 CGTGAATGCAACAAATAGAGATTTGTATGACACCCGAGCGGCTAATGGTTG 460  
Db 596 GGTACTACCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 655  
Qy 461 ATCTTCACTTACTCAAGAGAAATGCGCTGCTTGTATCATGT 509  
Db 656 TTTTMTATTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTT 704

RESULT 2 486 bp mRNA linear EST 22-JUL-2003  
CF088325  
LOCUS OHM18L03.yg.ab1 OH M sunflower H. argophyllus Helianthus argophyllus  
DEFINITION OHM18L03.yg.ab1 OH M sunflower H. argophyllus Helianthus argophyllus  
CDNA clone OHM18L03, mRNA sequence.

ACCESSION CF088325.1 GI:33127392  
VERSION CF088325.1  
KEYWORDS EST.  
SOURCE Helianthus argophyllus  
ORGANISM Helianthus argophyllus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; campanulids; Asterales; Asteraceae; Helianthaceae;  
Heliantheae; Helianthus.

REFERENCE 1 (bases 1 to 486)  
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
Lin, H., Van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,  
Ellison, P., Kolman, J., Slabough, M.S., Livingston, K., Zhou, Y.,  
Lai, Z., Church, S., Jackson, L. and Bradford, K.  
Letuce and Sunflower ESTs from the Composite Genome Project  
http://compgenome.ucdavis.edu/  
Unpublished (2002)

JOURNAL TITLE  
COMMENT Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Aremundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]  
belongs to contig OH\_CA\_Contig2435\_2, see http://cgdb.ucdavis.edu/  
for details.

FEATURES  
SOURCE Location/Qualifiers  
1..486

/organism="Helianthus argophyllus"  
/mol\_type="mRNA"  
/db\_xref="taxon:73275"  
/clone="OHM18L03"  
/lab\_host="E. coli"  
/clone\_lib="OH M sunflower H. argophyllus"  
/note="Vector: pRCRNASFIAB; The library was constructed  
from three different sources (seedling, root and leaf) of  
RNA from a single genotype. cDNAs were pooled and  
directionally cloned into a custom medium-copy vector.

Details of library construction can be obtained at  
http://cgdb.ucdavis.edu/

Query Match 5.1%; Score 41; DB 14; Length 486;  
Best Local Similarity 47.8%; Pred. No. 0.53;  
Matches 119; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

Qy 460 GATCTTCACTTACTTACCAAGAAATGCGCTGCTTGTATGATGTCGGCAAT 519  
Db 62 GATACACATCTTAATCAAGAGAGATGACCTCATATGTGAACAGTCACAGGT 121  
Qy 520 CAGATGCGTGGCAAAATTAACGCGCTGCTGTAATACATCCGTGATATTTC 579  
Db 122 AAACCTGTCGCGCAGAGCTATGTCGGGAAAGAGAGATTGTGCTGAACATATTGA 181  
Qy 580 TTAGAAAAGATTTTATGTCGCGGATTTATTTGATGCTTACCGCAATTTGGCGTGA 639  
Db 182 AAGAAAACGGGAAGTTAATGCGCAACATATGCAATGAACAGCATCCCGAGATA 241  
Qy 640 GTTATGCGATGACCATTAACGCGCTGATCAATACCGGGTATCGTACATCAAT 699  
Db 242 TACATTTGCGATTAAGTTCTTGTGTGCTAGACAAAGAGGATTCGTGAGTACCCCT 301  
Qy 700 AAAATGGCC 708  
Db 302 ACCGTGGC 310

RESULT 3 746 bp DNA linear GSS 29-JUL-2003  
CC876255  
LOCUS ZMNBBD0194M17.x ZMNBBD Zea mays subsp. mays genomic clone  
DEFINITION ZMNBBD0194M17.x ZMNBBD Zea mays subsp. mays genomic clone  
CDNA clone ZMNBBD0194M17, genomic survey sequence.

ACCESSION CC876255.1 GI:3306183  
VERSION CC876255.1  
KEYWORDS GSS.  
SOURCE Zea mays subsp. mays (maize)  
ORGANISM Zea mays subsp. mays (maize)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 746)  
Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.,  
and Wing, R.  
Sequencing of the maize genome  
Unpublished (2003)  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu

JOURNAL TITLE  
COMMENT FORWARD: T7  
BACKWARD: M13r  
Plate: 0194 row: M column: 17  
Seq primer: M13r  
Class: BAC ends.

FEATURES  
SOURCE Location/Qualifiers  
1..746

/organism="Zea mays subsp. mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/sub\_species="mays"  
/db\_xref="taxon:4578"  
/clone="ZMNBBD0194M17"  
/lab\_host="DH10B"  
/clone\_lib="ZMNBBD"  
/note="Vector: pReloBAC11; Site\_1: HindIII; Site\_2:  
HindIII; Zea mays L. ssp. mays"

ORIGIN

Query Match 5.1%; Score 40.8; DB 29; Length 746;  
 Best Local Similarity 59.5%; Pred. No. 0.74;  
 Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 388 GCTTATTAACCCGGAATGTCACAAAGAAATGTTGGATCGACCCGAGCGC 447  
 DB 51 GCTTATTAACCTCTAGGTTCTCAAGCAGATGATGATCTGAATCTGGCCAAAGTTC 110  
 QY 448 GCTAATGTTGATCTTCACTTACTCAAGAAGAAATGCCGCTGCTTTGA 503  
 DB 111 CATACTGATGACGCTACTGCAATTTAGCACTAATCAATGCAATGCGCAATTAA 166

RESULT 4  
 B2723869 826 bp DNA linear GSS 24-FEB-2003  
 LOCUS PUBS097D ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMBMTA097A17,  
 DEFINITION genomic survey sequence.  
 ACCESSION B2723869 GI:28519534  
 VERSION B2723869.1 GI:28519534  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 826)  
 Whiteley,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,  
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
 Bennetzen,J.  
 Maize Genomics Consortium  
 Unpublished (2003)  
 CONTACT: Cathy Whiteley  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whiteley@tigr.org  
 Seq primer: TF  
 Classes: sheared ends.  
 Location/Qualifiers  
 1..826  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_11b="ZM097A17"  
 /note="Vector: PCR4-TOPO; Site\_1: EcoRI; 0.6-1.0 kb high  
 COT selected genomic DNA library"

ORIGIN

Query Match 5.1%; Score 40.8; DB 28; Length 826;  
 Best Local Similarity 59.5%; Pred. No. 0.78;  
 Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 388 GCTTATTAACCCGGAATGTCACAAAGAAATGTTGGATCGACCCGAGCGC 447  
 DB 227 GCTTATTAACCTCTAGGTTCTCAAGCAGATGATGATCTGAATCTGGCCAAAGTTC 286  
 QY 448 GCTAATGTTGATCTTCACTTACTCAAGAAGAAATGCCGCTGCTTTGA 503  
 DB 287 CATACTGATGACGCTACTGCAATTTAGCACTAATCAATGCAATGCGCAATTAA 342

RESULT 5  
 CC389867 919 bp DNA linear GSS 19-MAY-2003  
 LOCUS PUB1W66TD ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMBMTA262J11,  
 DEFINITION genomic survey sequence.  
 ACCESSION CC389867

VERSION  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 919)  
 Whiteley,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,  
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
 Bennetzen,J.  
 Maize Genomics Consortium  
 Unpublished (2003)  
 CONTACT: Cathy Whiteley  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whiteley@tigr.org  
 Seq primer: TF  
 Classes: sheared ends.  
 Location/Qualifiers  
 1..919  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_11b="ZM06110\_KB"  
 /note="Vector: PCR4-TOPO; Site\_1: EcoRI; 0.6-1.0 kb high  
 COT selected genomic DNA library"

ORIGIN

Query Match 5.1%; Score 40.8; DB 28; Length 919;  
 Best Local Similarity 59.5%; Pred. No. 0.82;  
 Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 388 GCTTATTAACCCGGAATGTCACAAAGAAATGTTGGATCGACCCGAGCGC 447  
 DB 506 GCTTATTAACCTCTAGGTTCTCAAGCAGATGATGATCTGAATCTGGCCAAAGTTC 565  
 QY 448 GCTAATGTTGATCTTCACTTACTCAAGAAGAAATGCCGCTGCTTTGA 503  
 DB 566 CATACTGATGACGCTACTGCAATTTAGCACTAATCAATGCAATGCGCAATTAA 621

RESULT 6  
 CF090249 438 bp mRNA linear EST 22-JUL-2003  
 LOCUS QHM3P15.Y9.dbl QH M sunflower H. argophyllus Helianthus argophyllus  
 DEFINITION cDNA clone QHM3P15, mRNA sequence.  
 ACCESSION CF090249  
 VERSION CF090249.1 GI:33129316  
 KEYWORDS EST.  
 SOURCE Helianthus argophyllus  
 ORGANISM Helianthus argophyllus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; campanulids; Asterales; Asteroideae; Asteraceae;  
 Heliantheae; Helianthus.  
 1 (bases 1 to 438)  
 Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,  
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,  
 Ellison,P., Kolman,U., Slabough,M.S., Livingston,K., Zhou,Y.,  
 Lai,Z., Church,S., Jackson,U. and Bradori,K.  
 lettuce and sunflower ESTs from the Composite Genome Project  
 http://compgenome.ucdavis.edu/  
 Unpublished (2002)  
 CONTACT: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Armadson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659  
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to config OH\_CA\_Config2435, see http://cspdb.ucdavis.edu/  
 for details.  
 Plate: OHM12 row: P column: 15.

## FEATURES

## source

Location/Qualifiers  
 1..438  
 /organism="Helianthus argophyllus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:73275"  
 /clone="OHM3P15"  
 /lab\_host="E.coli"  
 /clone\_1lb="OH M sunflower H. argophyllus"  
 /note="Vector: pBRCDNA5flab; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and directionally cloned into a custom medium-copy vector. Details of library construction can be obtained at http://cspdb.ucdavis.edu/"

## ORIGIN

Query Match 5.1%; Score 40.4; DB 14; Length 438;  
 Best Local Similarity 47.3%; Pred. No. 0.76;  
 Matches 122; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 451 AATTGTTGATCTTCATCTTACTCAAGAAAGAAATGCCGCTGCTTTGATCATGTC 510  
 DB 44 AAATCATATATACACATCTTAAATCAAGAAAGATGCTCAGATATGTAAGATC 103  
 QY 511 GCGCAATTCAGCATGGTGGCAAAATTAACGGCTGCGCTGTAATACATCCGTGTA 570  
 DB 104 CAACAAGGTAAATTCATGCGCAGATAGTTGGGAGGAGGCGACATGCTCTGTAAC 163  
 QY 571 CTATTTTCATTAGAAAAGATTTTAGTGGCGGATTTAGATGCCCAATTT 630  
 DB 164 ATATTTGAAAAGAAAACAAAGATTAATGCGCAATTTGCTAAGAAATATCATC 223  
 QY 631 GGCGTCAAGTTATATGCGATGACATTAAGCGCGCTTGATCAATCCGGGTGATGTC 690  
 DB 224 CTGAGATATATCATTTTGCATAGAGTTCTTGTTGGTGAACAAAGGGATTTGTCG 283  
 QY 691 AGTACCAATAAATGGCC 708  
 DB 284 AGTACCCTTACCCTTGGC 301

## RESULT 7

## CF087071

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

CF087071 464 bp mRNA linear EST 22-JUL-2003  
 OHM12N04.Y9.ab1 OH M sunflower H. argophyllus Helianthus argophyllus  
 cDNA clone OHM12N04, mRNA sequence.  
 CF087071  
 CF087071.1 GI:33126122  
 EST.  
 Helianthus argophyllus  
 Helianthus argophyllus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;  
 Heliantheae; Helianthus.  
 1 (bases 1 to 464)  
 Kozik, A., Michelmore, R.W., Kapp, S., Matvienko, M., Rieseberg, L.,  
 Lin, H., Van Damme, W., Davelle, D., Chevalier, P., Ziegler, J.,  
 Ellisen, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,  
 Lai, Z., Church, S., Jackson, L. and Bradford, K.  
 Lecture and Sunflower ESTs from the Composite Genome Project  
 http://compgenome.ucdavis.edu/  
 unpublished (2002)  
 Contact: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Armadillo Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659  
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to config OH\_CA\_Config2435, see http://cspdb.ucdavis.edu/  
 for details.  
 Plate: OHM12 row: N column: 04.

## FEATURES

## source

Location/Qualifiers  
 1..464  
 /organism="Helianthus argophyllus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:73275"  
 /clone="OHM12N04"  
 /lab\_host="E.coli"  
 /clone\_1lb="OH M sunflower H. argophyllus"  
 /note="Vector: pBRCDNA5flab; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and directionally cloned into a custom medium-copy vector. Details of library construction can be obtained at http://cspdb.ucdavis.edu/"

## ORIGIN

Query Match 5.1%; Score 40.4; DB 14; Length 464;  
 Best Local Similarity 47.3%; Pred. No. 0.78;  
 Matches 122; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 451 AATTGTTGATCTTCATCTTACTCAAGAAAGAAATGCCGCTGCTTTGATCATGTC 510  
 DB 53 AAATCATATATACACATCTTAAATCAAGAAAGATGCTCAGATATGTAAGATC 112  
 QY 511 GCGCAATTCAGCATGGTGGCAAAATTAACGGCTGCGCTGTAATACATCCGTGTA 570  
 DB 113 CAACAAGGTAAATTCATGCGCAGATAGTTGGGAGGAGGCGACATGCTCTGTAAC 172  
 QY 571 CTATTTTCATTAGAAAAGATTTTAGTGGCGGATTTAGATGCCCAATTT 630  
 DB 173 ATATTTGAAAAGAAAACAAAGATTAATGCGCAATTTGCTAAGAAATATCATC 232  
 QY 631 GGCGTCAAGTTATATGCGATGACATTAAGCGCGCTTGATCAATCCGGGTGATGTC 690  
 DB 223 CTGAGATATATCATTTTGCATAGAGTTCTTGTTGGTGAACAAAGGGATTTGTCG 292  
 QY 691 AGTACCAATAAATGGCC 708  
 DB 293 AGTACCCTTACCCTTGGC 310

## RESULT 8

## AZ396257/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

AZ396257 638 bp DNA linear GSS 03-OCT-2000  
 M0160C17R Mouse 10kb plasmid U96C1M library Mus musculus genomic  
 clone U96C1M0160C17 R, genomic survey sequence.  
 AZ396257  
 AZ396257.1 GI:10511329  
 GSS.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 638)  
 Dunn, D., Aoyagi, A., Barber, M., Baecorn, T., Duval, B., Hamill, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von  
 Niederhausern, A. and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plates: 0168 row: C column: 17  
 Seq primer: CACACGAGAAACAGCATACC  
 Class: plasmid ends  
 High quality sequence stop: 638.  
 Location/Qualifiers

## FEATURES

source

1. 638  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUC1M0160C17"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_1lb="Mouse 10kb plasmid UUC1M library"  
 /note="Vector: pMD42m; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/Anares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi:4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

## Query Match

Best Local Similarity 52.7%; Pred. No. 1;  
 Matches 87; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 411 CAACAATAAGAAATTTTGTATGACCCGAGCGGCTAATGTTGATCTTCAATAC 470  
 DB 622 CAGATGAAAGAAATTTTGGCGCTATCCACAACCTGCTTACTTAAGAAACAATA 563  
 QY 471 TTACTCAAGAAATGCGCTGCTTTGATCATGTCCGCAATTCAGCATGCGTG 530  
 DB 562 ATTGTCATATTAAGAAAGCAGAACCTTTCTTCTGTGTCGAGAAATTAAGTGAA 503  
 QY 531 GCAAAATTAACCGCTGCGGTGATACATCCCTGTACTATT 575  
 DB 502 GAGAAATTAACACCTCATAGAAATCACTCATGTCAGAGATT 458

RESULT 9  
 AJ539921 401 bp mRNA linear EST 12-FEB-2003  
 LOCUS AJ539921 Hasens3 Helianthus annuus cDNA clone HM007, mRNA  
 DEFINITION

ACCESSION AJ539921  
 VERSION AJ539921  
 KEYWORDS GI:28370224  
 SOURCE EST.

ORGANISM Helianthus annuus (common sunflower)  
 Helianthus annuus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;  
 Heliantheae; Helianthus.

REFERENCE 1 (bases 1 to 401)  
 Tamborindesguay, C., Liboz, T., Petitprez, M. and Genzribitel, L.

AAn expressed-sequence-tag database of the sunflower protoplast  
 Unpublished (2003)  
 JOURNAL Contact: Tamborindesguay C

Laboratoire de Biotechnologie et Amélioration des Plantes  
 Institut National Polytechnique de Toulouse - Ecole National  
 Supérieure Agronomique de Toulouse  
 IRR40, Pole de Biotechnologie Vegetale, 18 chemin de Borde Rouge,  
 Auzeville, CASTANER TOLOSAN 31326, France.  
 Location/Qualifiers

## FEATURES

source

1. 401  
 /organism="Helianthus annuus"  
 /mol\_type="mRNA"  
 /culturivar="Em11"  
 /db\_xref="taxon:4232"  
 /clone="HM0007"  
 /cissue\_type="hypocotyl"  
 /cell\_type="protoplast"  
 /dev\_stage="1- to 5-days old protoplast"  
 /clone\_1lb="Hasens3"

## ORIGIN

Query Match 5.0%; Score 39.8; DB 9; Length 401;  
 Best Local Similarity 47.4%; Pred. No. 1.1;  
 Matches 119; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 458 TTGATCTTCACTTACTTACTCAAGAAATGCGGCTGCTTTGATCATGTGCGCAA 517  
 DB 33 TAGATCACATTACTTAATCAAGAGATGCGCTCAGTATGTGAACAAGTCAACAG 92  
 QY 518 TTGATGATGCGGCAAAATTAAGCGGCTGCGGTGAATACATCCGCTACTATTT 577  
 DB 93 GTAAACTCATGTGCGCAAGTTAGTGGGAAAGAGACATGCTGTAAACATATTG 152  
 QY 578 CATTAGAAAGAGTTTATGTCGCGGATTTATTTGATGCTTACGCCAAATTTGGCGTCG 637  
 DB 153 GAAAGGAAACACAAAGTTAATGCCCAATTTCTGAGGAACATCATCATCAGCA 212  
 QY 638 AAGTTATGCGATGACCATTAAGCGCGCTGATCAATCCGGGATGATGATGACCA 697  
 DB 213 TTTCATTTGCGATGAGTTCTTGTGTGGTGAACAAAGGGATTTGCTAGATGCC 272  
 QY 698 ATAAATGCGC 708  
 DB 273 CTACTGTGGC 283

## RESULT 10

CF090254

LOCUS

DEFINITION

CF090254 435 bp mRNA linear EST 22-JUL-2003  
 OMM3P22.YG.ab1 QH M sunflower H. argophyllus Helianthus argophyllus  
 cDNA clone OMM3P22, mRNA sequence.

ACCESSION CF090254  
 VERSION CF090254  
 KEYWORDS GI:33129321  
 SOURCE EST.

ORGANISM Helianthus argophyllus  
 Helianthus argophyllus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;  
 Heliantheae; Helianthus.

REFERENCE 1 (bases 1 to 435)  
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
 Jin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,  
 Ellison, P., Kolkmann, D., Slabaugh, M.S., Livingston, K., Zhou, Y.,  
 Lai, Z., Church, S., Jackson, L. and Bradford, K.

lettuce and sunflower ESTs from the Compositeae Genome Project  
 http://compgenome.ucdavis.edu/  
 Unpublished (2002)

CONTACT: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore lab  
 University of California at Davis (UCD)  
 Amundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659

Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]  
 belongs to config\_QH\_CA\_Config435, see http://cgpdb.ucdavis.edu/

for details.

Place: QM3 row: P column: 22.

## FEATURES

Location/Qualifiers

1..435

/organism="Helianthus argophyllus"

/mol\_type="mRNA"

/db\_xref="taxon:73275"

/clone="QM3P22"

/lab\_host="E.coli"

/clone\_lib="QH M sunflower H: argophyllus"

/note="Vector: pBRCDNA5flab; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and directionally cloned into a custom medium-copy vector. Details of library construction can be obtained at <http://cgdb.ucdavis.edu/>"

## ORIGIN

Query Match

Best Local Similarity 47.4%; Pred. No. 1.2; Length 435;

Matches 119; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 458 TTGATCTTCACTTACTTACTTAAAGAAAGATCCGCTGCTTGTATCATGTGCGCAAA 517  
 DB 27 TAGATACACACATCTAATCAACAAAGAGATGCGCTCATGTATGTGAACAGTCCAAAG 86  
 QY 518 TTGAGCATGCGTGGCAAAATTAACGCGCTGCGGTGATACATCCGCTGACTATTTT 577  
 DB 87 GTAAACCTTGCTGGCGAGATTAGTGGGAGAGAGCGACTGTGCTGTAAACATTAATTG 146  
 QY 578 CATTAGAAAAGATTGTTAGTGGCGATATTATGATGCTTACGCGCAAAATTTGCGCTGC 637  
 DB 147 AAAGAGAAAACAAAGATTAATGCCCAATTTTCTTAAGAACTATCATCTCGAGA 206  
 QY 638 AACTTAAATGCCAGACCATTAACGCCGCTGTATCAATCCGGGGTGAATGCTAGTACA 697  
 DB 207 TATACATTGCGATGAGTTCTTGTGTGGTGAACAAAGGAGATTGCTGTAGTACC 266  
 QY 698 ATAAATGCGC 708  
 DB 267 CTACCGTTGCG 277

## RESULT 11

BX500256/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

437 bp mRNA linear EST 04-SEP-2003  
 DXFZ779J2254 r1.779 (synonym: hnccl) Homo sapiens cDNA clone  
 DXFZ779J2254 5', mRNA sequence.  
 BX500256  
 BX500256.1 GI:32019447  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 437)  
 Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., Oesinger, A.,  
 Fodor, G., Han, M. and Wiemann, S.  
 EST (Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., et al.)  
 Unpublished (2003)  
 Contact: MIPS  
 MIPS  
 Ingolstaedter Landstr. 1, D-85764 Neuberg, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;  
 sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
 Braunschweig/Germany) within the cDNA sequencing consortium of the  
 German Genome Project.  
 No 5' sequence available.  
 This clone (DXFZ779J2254) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: [clone@rzpd.de](mailto:clone@rzpd.de).

## FEATURES

Location/Qualifiers

1..437

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="DXFZ779J2254"

/tissue="liver"

/dev\_stage="fetal"

/lab\_host="MD10B"

/clone\_lib="779 (synonym: hnccl)"

/note="Vector: pSPORT1\_Sfi, Site\_1: SfiIA, Site\_2: SfiIB"

## ORIGIN

Query Match

Best Local Similarity 53.2%; Pred. No. 1.3; Length 437;

Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 237 TGGAGAACCTGTCGATGAGGATTTTGTATGACGAGACGATATTGTCGCAAAAGT 296  
 DB 275 TCGAGACCTTTCATGTCATTTTAAAGCTGTGAGATTTGTTTCATTAAGT 216  
 QY 297 CCATCTTATCCATATTTCATGCGCATCCGCTGTTGATGGAACCCCAAGCGATCC 356  
 DB 215 CTGCTTTTCAAGTGAACATCACTCTCCGCTTTGTGTGCTTAAATGATGT 156  
 QY 357 GAATGTTGAGATGTCAGTATTTCCATTAAGCTTTAT 394  
 DB 155 GACTTTTTCAGAGACCTTTCATTTCTGTACATTTGAT 118

## RESULT 12

CF091479

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

447 bp mRNA linear EST 22-JUL-2003  
 OHW8P22.Y9.ab1 QH M sunflower H: argophyllus Helianthus argophyllus  
 cDNA clone QH8P22, mRNA sequence.  
 CF091479  
 CF091479.1 GI:33130546  
 EST.  
 Helianthus argophyllus  
 Helianthus argophyllus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; campanulids; Asterales; Asterales; Asteraceae; Asteroideae;  
 Heliantheae; Helianthus.  
 1 (bases 1 to 447)  
 Kozik, A., Michelmore, R.W., Krapp, S., Matvienko, M., Rieseberg, L.,  
 Liu, H., Van Damme, J., Lavelle, D., Chevalier, P., Ziegler, J.,  
 Ellison, P., Kolkmann, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,  
 Lai, Z., Church, S., Jackson, L. and Bradford, K.  
 Lettuce and Sunflower ESTs from the Compositeae Genome Project  
<http://compenomics.ucdavis.edu/>  
 Unpublished (2002)  
 Contact: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmndson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: [akozik@ucdavis.org](mailto:akozik@ucdavis.org) [[michelmorevegmail.ucdavis.edu](mailto:michelmorevegmail.ucdavis.edu)]  
 belongs to contig QH\_CA\_Contig2435\_2, see <http://cgdb.ucdavis.edu/>  
 for details.  
 Plate: QH8 row: F column: 22.  
 Location/Qualifiers  
 1..447  
 /organism="Helianthus argophyllus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:73275"  
 /clone="QH8P22"  
 /lab\_host="E.coli"  
 /clone\_lib="QH M sunflower H: argophyllus"  
 /note="Vector: pBRCDNA5flab; The library was constructed from three different sources (seedling root and leaf) of RNA from a single genotype. cDNAs were pooled and

directionally cloned into a custom medium-copy vector.  
Details of library construction can be obtained at  
<http://cspdb.ucdavis.edu/>

directionally cloned into a custom medium-copy vector.  
Details of library construction can be obtained at  
<http://cspdb.ucdavis.edu/>

## ORIGIN

Query Match 4.9%; Score 39.4; DB 14; Length 447;  
Best Local Similarity 47.4%; Pred. No. 1.6;  
Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Query Match 4.9%; Score 39.4; DB 14; Length 454;  
Best Local Similarity 47.4%; Pred. No. 1.6;  
Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 460 GATCTTCATCTTCTCTCAAGAGAAATGCGCGCTTTTGATCATGTCCGCAATT 519

QY 460 GATCTTCATCTTCTCTCAAGAGAAATGCGCGCTTTTGATCATGTCCGCAATT 519

DB 21 GATACACACATCTTAATACAAAGAGATGGCTTCAGATGTGAACAGTCAACAGT 80

DB 57 GATACACACATCTTAATACAAAGAGATGGCTTCAGATGTGAACAGTCAACAGT 116

QY 520 CAGCATGCGTGGCAAAATTAACGCGTGGTGAATACATCCGCTGACTATTTC 579

QY 520 CAGCATGCGTGGCAAAATTAACGCGTGGTGAATACATCCGCTGACTATTTC 579

DB 81 AAACTCTGTCGCGCAAGCTAGTCGGGAAGAGATGTGCTGTAAACATTAATGA 140

DB 117 AAACTCTGTCGCGCAAGCTAGTCGGGAAGAGATGTGCTGTAAACATTAATGA 176

QY 580 TTGAAAAAGAGTTTATGTCGGATATTATTGATGCTTACGCCAAATTTGGCTGAA 639

QY 580 TTGAAAAAGAGTTTATGTCGGATATTATTGATGCTTACGCCAAATTTGGCTGAA 639

DB 141 AAGGAAAACGCGAAGGTTAATGCGCAAACTATCGCAAGAACGACATCCCGAGATA 200

DB 177 AAGGAAAACGCGAAGGTTAATGCGCAAACTATCGCAAGAACGACATCCCGAGATA 236

QY 640 GTTATCGCATGACATTAACGCGCTGTGATCAATCCGGGGTATGTCATGACCAAT 699

QY 640 GTTATCGCATGACATTAACGCGCTGTGATCAATCCGGGGTATGTCATGACCAAT 699

DB 201 TACATTGGCATGAGTTCTTGTGTGGTGAACAAAGAGGATTTGTGTGATGATCCCT 260

DB 237 TACATTGGCATGAGTTCTTGTGTGGTGAACAAAGAGGATTTGTGTGATGATCCCT 296

QY 700 AAAATGCCC 708

QY 700 AAAATGCCC 708

DB 261 ACCGTTGGC 269

DB 297 ACCGTTGAC 305

## RESULT 13

## RESULT 14

CF090121

CF096847

LOCUS QHM3J13.YG.ab1 OH M sunflower H. argophyllus linear EST 22-JUL-2003

LOCUS QHM2F11.YG.ab1 OH N sunflower H. argophyllus (drought stress) linear EST 22-JUL-2003

DEFINITION CDMN clone QHM3J13, mRNA sequence.

DEFINITION HElianthus argophyllus CDMN clone QHM2F11, mRNA sequence.

ACCESSION CF090121

ACCESSION CF096847

VERSION CF090121.1 GI:33129188

VERSION CF096847.1 GI:33135914

## KEYWORDS

## KEYWORDS

## SOURCE

## SOURCE

## ORGANISM

## ORGANISM

## EST

## EST

## REFERENCE

## REFERENCE

## AUTHORS

## AUTHORS

## TITLE

## TITLE

## JOURNAL

## JOURNAL

## COMMENT

## COMMENT

Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Amundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-752-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]  
belongs to contig OH\_CA\_Config2435\_2, see <http://cspdb.ucdavis.edu/>  
for details.

Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Amundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-752-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]  
belongs to contig OH\_CA\_Config2435\_2, see <http://cspdb.ucdavis.edu/>  
for details.

## FEATURES

## FEATURES

source

source

1..454  
Location/Qualifiers

1..401  
Location/Qualifiers

/organism="Helianthus argophyllus"

/organism="Helianthus argophyllus"

/mol\_type="mRNA"

/mol\_type="mRNA"

/db\_xref="taxon:73275"

/db\_xref="taxon:73275"

/clone="QHM3J13"

/clone="QHM2F11"

/lab\_host="E.coli"

/lab\_host="E.coli"

/note="Vector: pBRCDNA51AB; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and

/note="Vector: pGEM-T; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and



## ORIGIN

a high-copy vector pGBM-T. Details of library construction can be obtained at <http://cspdb.ucdavis.edu/>

Query Match 4.9%; Score 39; DB 14; Length 401;  
Best Local Similarity 46.5%; Pred. No. 1.9;  
Matches 126; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 455 GGTGATCTTCACTTACTTACTCAAGAAGAAATGCCCGCTTGTGATCATGTGCGC 514  
DB 5 GGATGACACACACATCTTAATCAAGAGATGCTCATATGTAACAAGTCCAC 64  
QY 515 AATTCAGATGCGTGGCAAAATTAACGCGCTGCTGAATACATCCGCTGACTAT 574  
DB 65 AAGGTAAACTCTGTCGCGAGCTGTGCGGAAAGAGATGTGCTGTAACATTA 124  
QY 575 TTTCATTAAGAAAAGATTTTATGTCGCGATTTATGATGCTACGCCAATTGGCG 634  
DB 125 TTGAAGAAGAAAACGGAAGTTAATGCGCAACTATGCGAAGAAACCATCCCG 184  
QY 635 TCGAAGTTAATGCGCATGACATTAAACGCCGCTGATCAATACCGGGGATTCGTA 694  
DB 185 AATATATACATTTGCGATAGATTTCTGTGTGCTGAACAGAGGGATGTGTGAGTA 244  
QY 695 CCAATTAATGCGCGCATCTTGTAAAGCAA 725  
DB 245 CCCCTACGCTTGCTAATAAAAAAAAAACA 275

## RESULT 15

AJ540178 451 bp mRNA linear EST 12-FEB-2003  
AJ540178 Helianthus annuus cDNA clone HM0013, mRNA  
DEFINITION sequence.

ACCESSION AJ540178  
VERSION AJ540178.1 GI:28370481  
KEYWORDS EST  
SOURCE Helianthus annuus (common sunflower)  
ORGANISM Helianthus annuus

REFERENCE  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
TITLE Spermatophyta; Magnoliopsida; core eudicots;  
JOURNAL asterids; campanulids; Asterales; Asteraceae; Asteroideae;  
COMMENT Heliantheae; Helianthus.  
1 (bases 1 to 451)  
Tamborindeguy, C., Liboz, T., Petitprez, M. and Gentschitzel, L.  
An expressed-sequence-tag database of the sunflower protoplast  
Unpublished (2003)  
Contact: Tamborindeguy C  
Laboratoire de Biotechnologie et Amélioration des Plantes  
Institut National Polytechnique de Toulouse - Ecole National  
Supérieure Agronomique de Toulouse  
IFR40, Pole de Biotechnologie Vegetale, 18 chemin de Borde Rouge,  
Auzerville, CASTANET TOLOSAN 31326, France.  
Location/Qualifiers

## FEATURES

source  
1..451  
/organism="Helianthus annuus"  
/mol\_type="mRNA"  
/cultivar="Emil"  
/db\_xref="taxon:4232"  
/clone="HM0013"  
/issue\_type="hyocoty1"  
/cell\_type="protoplast"  
/dev\_stage="1- to 5-days old protoplast"  
/clone\_lib="Hsems3"

## ORIGIN

Query Match 4.9%; Score 39; DB 9; Length 451;  
Best Local Similarity 47.7%; Pred. No. 2.1;  
Matches 114; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 458 TTGATCTTCACTTACTTACTCAAGAAGAAATGCCCGCTTGTGATCATGTGCGCAA 517  
DB 32 TAGATACACACATCTTAATCAAGAGAGTGGCCCATGATGTAACAAGTCCAAAG 91

QY 518 TTGACATGCGTGGCAAAATTAACGCGCTGCGTTGAATACATCCGTAATTTT 577  
DB 92 GTAAACTTGTGCGCAGAGCTAGTGGGGAAGAGATTGTGTGTAACATPATG 151  
QY 578 CATTAGAAAAGATTTTATGTCGCGAATTAATGATCCTACGCCAATTGGCGTCG 637  
DB 152 AAAGGAAAACGGAGGTTAATGCGCAACTATGCAAGAGAACACGATCCCGAGA 211  
QY 638 AAGTTATCGCATGACCATTAACGCCGCTTGATCAATACGGGGTGAATCGTCA 696  
DB 212 TATACATTTGCGATAGATTTCTGTGTGCGGTGAACAAAGGGGATTGTGTGATAC 270

Search completed: April 30, 2004, 18:35:03  
Job time : 2273.3 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:32 ; Search time 93.7866 Seconds

(without alignments)  
4721.887 Million cell updates/sec

Title: US-10-603-260-2

Perfect score: 798  
Sequence: 1 agctgttcacacaaagctcta.....ttcaacctgtttagaggt 798

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCITUS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.6	4.5	552	4	US-09-107-532A-286
2	35.2	4.4	4392	1	US-08-026-138E-5
3	34	4.3	1326	4	US-09-489-039A-2920
4	33.6	4.2	2628	4	US-09-294-531B-5
5	33.4	4.2	474	4	US-09-621-976-18033
6	33.4	4.1	1830121	4	US-09-557-884-1
7	33	4.1	1830121	4	US-09-643-990A-1
8	32.6	4.1	552	4	US-09-489-039A-10
9	32.6	4.1	825	4	US-09-134-001C-2442
10	32.4	4.1	696	3	US-08-998-41E-1134
11	32.4	4.1	1620	3	US-08-925-230-2
12	32.4	4.1	1620	3	US-09-712-372-2
13	32.4	4.1	11466	4	US-08-956-171E-444
14	32.4	4.0	846	4	US-09-328-352-4044
15	32	4.0	7493	1	US-08-212-133A-7
16	32	4.0	7493	1	US-08-474-503-5
17	32	4.0	7493	2	US-08-670-707A-5
18	32	4.0	7493	3	US-09-037-601-5
19	32	4.0	7493	3	US-09-315-179-5
20	32	4.0	7493	5	PCT-US94-13200-5
21	31.6	4.0	378	3	US-09-328-111-191
22	31.6	4.0	1446	2	US-08-933-750C-91
23	31.6	4.0	1724	3	US-09-234-613-91
24	31.6	4.0	1724	3	US-09-620-312D-650
25	31.4	3.9	1321	4	US-09-620-312D-790
26	31.4	3.9	1462	4	US-09-620-312D-788
27	31.4	3.9	1519	4	US-09-620-312D-789

28	31.4	3.9	2758	4	US-09-221-017B-441	Sequence 441, App
29	31.2	3.9	2290	4	US-09-620-312D-891	Sequence 891, App
30	31	3.9	477	4	US-09-489-039A-801	Sequence 801, App
31	31	3.9	1089	4	US-09-134-000C-901	Sequence 901, App
32	30.8	3.9	648	4	US-09-543-681A-3313	Sequence 3313, App
33	30.8	3.9	1152	4	US-09-540-236-1145	Sequence 1145, App
34	30.8	3.9	119211	4	US-09-596-002-40	Sequence 40, App
35	30.6	3.8	500	3	US-09-141-000-2	Sequence 2, App
36	30.6	3.8	678	3	US-09-232-479-19	Sequence 19, App
37	30.6	3.8	754	4	US-09-784-990-19	Sequence 19, App
38	30.6	3.8	754	4	US-09-221-017B-941	Sequence 941, App
39	30.6	3.8	2829	3	US-08-651-843A-53	Sequence 53, App
40	30.6	3.8	2829	3	US-08-974-549A-220	Sequence 220, App
41	30.6	3.8	2829	3	US-08-854-050-53	Sequence 53, App
42	30.6	3.8	2829	4	US-09-430-123-53	Sequence 53, App
43	30.6	3.8	2829	4	US-09-402-181E-220	Sequence 220, App
44	30.6	3.8	2829	4	US-09-721-456-220	Sequence 220, App
45	30.6	3.8	3300	4	US-09-268-347-31	Sequence 31, App

## ALIGNMENTS

RESULT 1  
US-09-107-532A-286  
Sequence 286, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 286:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 552 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: circular  
HYPOTHETICAL: NO (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...552





J. Craig Venter  
TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186P1C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1  
Query Match 4.1%; Score 33; DB 4; Length 1830121;  
Best Local Similarity 65.8%; Pred. No. 52;  
Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
CY 586 AAGAGTTTATGAGCGGATATTATGCTAGCCAAATTGGCGTGAAGTTAAT 645  
DB 1807122 AACGATTTTGGTCGGCTGCAATCTTCAAGCCGTCGCAACTTTGCTTCGATTAAAC 1807181  
CY 646 CGCATGACCATTA 658  
DB 1807182 AGTAATGCCATCA 1807194  
RESULT 8  
US-09-489-039A-10/c  
Sequence 10, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 10  
LENGTH: 552  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae

US-09-489-039A-10  
Query Match 4.1%; Score 32.6; DB 4; Length 552;  
Best Local Similarity 57.3%; Pred. No. 0.52;  
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
CY 262 GATGACACGAGACGACTATTGTGCGCAAAAGTCATCTTATCTATTATTCACG 321  
DB 175 GTTTCAGCCACACGAGTGTGTAAGCCAGACAGATGCTTACGCACTTGAACGACG 116  
CY 322 GATCCGCTGTTGATGAGCAACCCAAACGGATCCGATGTT 364  
DB 115 GATGACACGATTCATTACTACCGCAACGCGCCGTCAGATT 73  
RESULT 9  
US-09-134-001C-2442/c  
Sequence 2442, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2442  
LENGTH: 825  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2442  
Query Match 4.1%; Score 32.6; DB 4; Length 825;  
Best Local Similarity 57.3%; Pred. No. 0.67;  
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
CY 65 TGTATTATGAGAACCATATGATGTCATCATATGATTTGTCGCTTGGCG 124  
DB 588 TATTAATTCGAAATTAATGTCGTAGCTTCGATCTAATATTTCTACTAAGTA 529  
CY 125 CGAATCTATCCAGTTTACTGCTGAACGCTCTAATCCAA 167  
DB 528 TCGAATGTTCCAGCTAATACAGCTGTGATGATATCAACAGA 486  
RESULT 10  
US-08-998-416-1134/c  
Sequence 1134, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8687  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 1134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 696 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1690UP  
US-08-998-416-1134

Query Match 4.1%; Score 32.4; DB 3; Length 696;  
Best Local Similarity 47.5%; Pred. No. 0.71;  
Matches 96; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 171 GCGAGATTGTGTAATGGCGCATTCCTGCGGAGATGATGACGAAGATATGACCGC 230  
DB 452 GCGACTATGCTGCTGCGGAGAGCTAGACAGCACTTTGCACTAGAGGTGCTGCGGATCA 393  
QY 231 TCATGTGTGAGAACCTGTGATGAGATTGATGACGAGACGCGTATTTGTGCGCA 290  
DB 392 ACTTAGGGGAGACGACTTCGAGAGGATAGAGCGCAAGCGGCGCTAGCGCGAG 333  
QY 291 AAAAGCCCTACTTATCCCTAATTTATGCGATCCGCTGCTGATGCGCAACCCCAAG 350  
DB 332 CAAGGGCGAAGCAGTACAAAGTATGCTGCGAGCGGAGCTAGAGTGACGCGCA 273  
QY 351 CGATCCGATGCTTGGAGTGTG 372  
DB 272 CGTGAAGATGAGGCGCGCTC 251

RESULT 11  
US-08-925-230-2  
Sequence 2, Application US/08925230  
Patent No. 6147194  
GENERAL INFORMATION:  
APPLICANT: Collart, Frank  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO IMPDH  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Arnold White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/925,230  
FILING DATE: September 8, 1997  
CLASSIFICATION: 536  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 5,665,583  
FILING DATE: 12-AUG-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: ARCD:274  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1620 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-925-230-2

Query Match 4.1%; Score 32.4; DB 3; Length 1620;  
Best Local Similarity 58.2%; Pred. No. 1.2;  
Matches 57; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 598 GTGGCGGATATTATGATGCTACGCCAATTGGCGTCCAGTATATCGCATGACCAT 657  
DB 404 GTGAGGATGTTTGTGAGCCAAAGCAGGCAATGCTCTGTGTATCCCATCAGCAT 463  
QY 658 AAACGCCCTTGATCAATACCGGGTGATGCTGATAC 695  
DB 464 ACAGCCCGATGGGAGTGCATGCTGCGGATCATTTTC 501

RESULT 12  
US-08-712-372-2  
Sequence 2, Application US/09712372  
Patent No. 6479628  
GENERAL INFORMATION:  
APPLICANT: Huberman, Eliezer  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO IMPDH  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/712,372  
FILING DATE: 13-NO. 6479628-2000  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/925,230  
FILING DATE: September 8, 1997  
APPLICATION NUMBER: US 5,665,583  
FILING DATE: 12-AUG-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: ARCD:274  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 474-7577  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1620 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-712-372-2

Query Match 4.1%; Score 32.4; DB 4; Length 1620;  
Best Local Similarity 58.2%; Pred. No. 1.2;  
Matches 57; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 598 GTGGCGATATATATGATGCTTACGCAATTTGGCGTGAAGTTATCGATGACCAT 657  
DB 404 GTGAGGATGCTTTTGAAGCAAGCCAGGATGCTTGTGATGCCATCAAGAT 463  
QY 658 AAAGCGCGCTTGATCAATACCGGGGTGATGTCAGTAC 695  
DB 464 ACAGCGCGATGGAGATGCTGACTGGCGCATCATTTTC 501

## RESULT 13

US-08-956-171E-444  
Sequence 444, Application US/08956171E  
Patent No. 6593114  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E  
FILING DATE: 20-Oct-1997  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861  
FILING DATE: January 5, 1996  
APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark U. Hyman  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB248P1

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 444:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1146 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

## SEQUENCE DESCRIPTION: SEQ ID NO: 444:

US-08-956-171E-444

Query Match 4.1%; Score 32.4; DB 4; Length 1146;  
Best Local Similarity 60.0%; Pred. No. 4;  
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 ATGTTTGTCACAAAGTCTTATTACATTGACCATCATGACATTAACCTTAATAAGC 60  
DB 11258 ATGTTGACATGATCACTTAATTTCTTGACACATATCATCTTAACAATATGCTTA 11317  
QY 61 CCGTTGTTATTAGGAGAGCCATTAGATT 90  
DB 11318 TCGTTGATTAGTAAGTATCAATTTTGTT 11347

## RESULT 14

US-09-328-352-4044/C  
Sequence 4044, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 4044  
LENGTH: 846  
TYPE: DNA

ORGANISM: Acinetobacter baumannii  
US-09-328-352-4044

Query Match 4.0%; Score 32; DB 4; Length 846;  
Best Local Similarity 47.9%; Pred. No. 1.1;  
Matches 92; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 367 AGTGTGATATTTTCCATTAAGCTTTTAAACCCGTGAATGTCAACCAATAGAGAT 426  
DB 601 AGCTTAATTTTACCTGACAGATCATTAATTCATGAGTAGAAGAAATAGCTGAC 542  
QY 427 TTGTGATCGAACCCGAGCGCGCTAATGTTGATTTTCACTTACTCAAGAGAA 486  
DB 541 CTGAACCTGAACCCGCGCTTACGCGATCAATTTGAGTTGAGTTTACCCAAATGTTA 482  
QY 487 ATGCGCGTGGCTTTGATCAATGCGCAATTCAGATCGTGGCAAAATTAACGGCT 546  
DB 481 AACAGTTTCTGCTTAATATGACTGTTAATTCACGCTCCGAGGACCAACCAATCAG 422  
QY 547 GCGGTTGAATAC 558  
DB 421 TCGCGTAATATAC 410

## RESULT 15

US-08-212-133A-7/C  
Sequence 7, Application US/08212133A  
Patent No. 5663060  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S.

TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kiplattick & Cody  
STREET: 100 Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/212,133A  
FILING DATE: March 11, 1994



```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type="terminal"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7471..7476
OTHER INFORMATION: /function="PolyA_signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="5'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 408..7367
OTHER INFORMATION: /product="Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschler, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 563060
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 7476
US-08-212-133A-7

Query Match          4.0%; Score 32; DB 1; Length 7493;
Best Local Similarity 62.5%; Pred. No. 4.2;
Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY      35 TCATCATGCACTTACTTAATAAATAGCCCGTGTATTATAGGAGACCATATGATGTCA 94
DB      895 TCTTTCAGGACTTGCACAAACATATAGTATGACTTTCACCGAGGAAACCTTATCATCTTCC 836
QY      95 CTATGATATGATTTGTCTG 114
DB      835 TTCCTCATTTGGCTTGTCTG 816
```

Search completed: April 30, 2004, 18:43:08  
Job time : 108.787 secs



Db 121 GGGCCGAATCTATCCAGGTTTACTGTGTAAGCCTCTAATCCAAATCGGCGAATGT 180  
QY 181 GGTAAATGGGCAATTCCTGCGGCGGAGTAGTATGACGAAGATATGACCGCTCATGTGGA 240  
Db 181 GGTAAATGGGCAATTCCTGCGGCGGAGTAGTATGACGAAGATATGACCGCTCATGTGGA 240  
QY 241 GAACCTGTGATGAGGATTTTATGATGAGGAGACGACGATTTTGTGCGGCAAAAGTCCAT 300  
Db 241 GAACCTGTGATGAGGATTTTATGATGAGGAGACGACGATTTTGTGCGGCAAAAGTCCAT 300  
QY 301 ACTTATCTAATTTTATTCAGCATCCGCTGTGTGATGGAACCCCAACCGGATCCGAT 360  
Db 301 ACTTATCTAATTTTATTCAGCATCCGCTGTGTGATGGAACCCCAACCGGATCCGAT 360  
QY 361 GGTGAGAGTCAATTTTCCCATTCAGCTTTATTTAAACCGGTGGAATGTCAAACAAATA 420  
Db 361 GGTGAGAGTCAATTTTCCCATTCAGCTTTATTTAAACCGGTGGAATGTCAAACAAATA 420  
QY 421 GAAGATTTTGTATGACACCCGAGCGGCTAATGTGATCTTATCTTACTCAA 480  
Db 421 GAAGATTTTGTATGACACCCGAGCGGCTAATGTGATCTTATCTTACTCAA 480  
QY 481 GAAAGAAATCCCGTGGCTTTTATGATGTCGCGCAAAATTCAGCATGCGTGGCAAAATTA 540  
Db 481 GAAAGAAATCCCGTGGCTTTTATGATGTCGCGCAAAATTCAGCATGCGTGGCAAAATTA 540  
QY 541 CGGCGTGGGTTGAATACACATCCGCTGACTAATTTTCAATGAAAAGATTTTATG 600  
Db 541 CGGCGTGGGTTGAATACACATCCGCTGACTAATTTTCAATGAAAAGATTTTATG 600  
QY 601 GCGGATATTATGATGCTACGCAAAATTTGGCGTCAAGTTTATCGCATGACCAATTA 660  
Db 601 GCGGATATTATGATGCTACGCAAAATTTGGCGTCAAGTTTATCGCATGACCAATTA 660  
QY 661 CGCGCTGTATCAATTCGCGGGGTGATCGTCAATCAATTAATGCGCGCATCTTTGTA 720  
Db 661 CGCGCTGTATCAATTCGCGGGGTGATCGTCAATCAATTAATGCGCGCATCTTTGTA 720  
QY 721 GCGAAGAGAGCAAAACGACGCTTATCGTCTTCCAGTCAATGAAGTCACTATTT 780  
Db 721 GCGAAGAGAGCAAAACGACGCTTATCGTCTTCCAGTCAATGAAGTCACTATTT 780  
QY 781 CAAACCTGTTTACGAGGT 798  
Db 781 CAAACCTGTTTACGAGGT 798

RESULT 2  
US-10-603-260-1  
Sequence 1, Application US/10603260  
Publication No. US20040009570A1  
GENERAL INFORMATION:  
APPLICANT: CJ Corporation  
TITLE OF INVENTION: An alkaline lipase from *Vibrio metchnikovii* RH530 and a  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/603,260  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Koparentin 1.71  
SEQ ID NO 1  
LENGTH: 2578  
TYPE: DNA  
ORGANISM: *Vibrio metchnikovii* RH530  
US-10-603-260-1

Query Match 100.0%; Score 798; DB 16; Length 2578;  
Best Local Similarity 100.0%; Pred. 2.9e-239;  
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTGTCAAAAGCTTATTTACATTTGACCATATCATGCACTTACCTTAAATTAAGC 60  
Db 703 ATGTTTGTCAAAAGCTTATTTACATTTGACCATATCATGCACTTACCTTAAATTAAGC 762

QY 61 CCGTGTGTTTATAGGAGCCATTATGATTTGCACTATGATATGATTTTGTCTGCGTCTT 120  
Db 763 CCGTGTGTTTATAGGAGCCATTATGATTTGCACTATGATATGATTTTGTCTGCGTCTT 822  
QY 121 GCGCGCAATCTATCCAGGTTTATCTGTGAAACGCTTATATCAATCGGCGCATGTGT 180  
Db 823 GCGCGCAATCTATCCAGGTTTATCTGTGAAACGCTTATATCAATCGGCGCATGTGT 882  
QY 181 GGTAAATGGGCAATTCCTGCGGCGGAGTAGTATGACGAAGATATGACCGCTCATGTGGA 240  
Db 883 GGTAAATGGGCAATTCCTGCGGCGGAGTAGTATGACGAAGATATGACCGCTCATGTGGA 942  
QY 241 GAACCTGTGATGAGGATTTTATGATGAGGAGACGACGATTTTGTGCGGCAAAAGTCCAT 300  
Db 943 GAACCTGTGATGAGGATTTTATGATGAGGAGACGACGATTTTGTGCGGCAAAAGTCCAT 1002  
QY 301 ACTTATCTAATTTTATTCAGCATCCGCTGTGTGATGGAACCCCAACCGGATCCGAT 360  
Db 1003 ACTTATCTAATTTTATTCAGCATCCGCTGTGTGATGGAACCCCAACCGGATCCGAT 1062  
QY 361 GGTGAGAGTCAATTTTCCCATTCAGCTTTATTTAAACCGGTGGAATGTCAAACAAATA 420  
Db 1063 GGTGAGAGTCAATTTTCCCATTCAGCTTTATTTAAACCGGTGGAATGTCAAACAAATA 1122  
QY 421 GAAGATTTTGTATGACACCCGAGCGGCTAATGTGATCTTATCTTACTCAA 480  
Db 1123 GAAGATTTTGTATGACACCCGAGCGGCTAATGTGATCTTATCTTACTCAA 1182  
QY 481 GAAAGAAATCCCGTGGCTTTTATGATGTCGCGCAAAATTCAGCATGCGTGGCAAAATTA 540  
Db 1183 GAAAGAAATCCCGTGGCTTTTATGATGTCGCGCAAAATTCAGCATGCGTGGCAAAATTA 1242  
QY 541 CGGCGTGGGTTGAATACACATCCGCTGACTAATTTTCAATGAAAAGATTTTATG 600  
Db 1243 CGGCGTGGGTTGAATACACATCCGCTGACTAATTTTCAATGAAAAGATTTTATG 1302  
QY 601 GCGGATATTATGATGCTACGCAAAATTTGGCGTCAAGTTTATCGCATGACCAATTA 660  
Db 1303 GCGGATATTATGATGCTACGCAAAATTTGGCGTCAAGTTTATCGCATGACCAATTA 1362  
QY 661 CGCGCTGTATCAATTCGCGGGGTGATCGTCAATCAATTAATGCGCGCATCTTTGTA 720  
Db 1363 CGCGCTGTATCAATTCGCGGGGTGATCGTCAATCAATTAATGCGCGCATCTTTGTA 1422  
QY 721 GCGAAGAGAGCAAAACGACGCTTATCGTCTTCCAGTCAATGAAGTCACTATTT 780  
Db 1423 GCGAAGAGAGCAAAACGACGCTTATCGTCTTCCAGTCAATGAAGTCACTATTT 1482  
QY 781 CAAACCTGTTTACGAGGT 798  
Db 1483 CAAACCTGTTTACGAGGT 1500

RESULT 3  
US-10-282-122A-41398  
Sequence 41398, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlson, Karl  
APPLICANT: Zyckind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: EUTRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patent version 3.1  
SEQ ID NO: 41398  
LENGTH: 717  
TYPE: DNA  
ORGANISM: *Vibrio cholerae*  
US-10-282-122A-41398

Query Match 50.9%; Score 405.8; DB 13; Length 717;  
Best Local Similarity 73.1%; Pred. No. 1.9e-116;  
Matches 521; Conservative 0; Mismatches 192; Indels 0; Gaps 0;  
QY 85 AAGATTCACATATGATATGATTTGCTGGCTTTGGCCGAATCTATCAGTTTAA 144  
DB 1 AAGATTCACATATGATATGATTTGCTGGCTTTGGCCGAATCTATCAGTTTAA 60  
QY 145 CTGATGAAACGCTATATCAATCGCCAGATTGTGTAATGGGCAATCCCTGCGGG 204  
DB 61 TTGATCAACGAGCAATCAACCGCCGATATGCGCTTTGGGCGATTCGGGTGT 120  
QY 205 ATATGATATGACGAATATGACCGCTCAATGTTGAGAGCACTCTGATGAGATTTGAT 264  
DB 121 TGGGTTTATGATGAGATATCTCAGCCACAAGGTGGCGAACCCTGACAAAGATTTGAT 180  
QY 265 GAGAGGAGCAGCATTTTGTGGGCAAAAGTCACTTATCTTAATTTTATCAGGAT 324  
DB 181 TCCGCGCTGCGCGCATCTGCGCGCAAAATCTTACTTACCCTTATCAGATGAT 240  
QY 325 CCGCTGTTGATGAGCAACCCCAACGCGATCGAATGTTGATGATGATTTCCCAT 384  
DB 241 CTTTGTGTCGATATCTTAAACGCGATCCAAAGGTTGAGATCATGATTTTCTAC 300  
QY 385 TACGCTTATTAACCCCGGAATCTCAACAAATAGAAATTTGTATGACACCCGAG 444  
DB 301 TATGCCCTGCTCAATCATTCACACGTCAACGATTTGAAGAGGCTGAGATGATTAAGAC 360  
QY 445 CCGCTAATTTGTTGATCTTCTATCTTACTCAAGAAGAAATGCGCTGCTTTGAT 504  
DB 361 CGTGTATTTGTTGATCTTCTATCTTACTCAAGAAGAAATGCGCTGCTTTGAT 420  
QY 505 CATGTCGCGCAATTCAGATGCGTGGCAAAATTAACCGCTGCGTTGATTAACATCC 564  
DB 421 CATGTCGCGCAATTCAGATGCGTGGCAAAATTAACCGCTGCGTTGATTAACATCC 480  
QY 565 GTGATCTATTTCTTATGAAAAAGTTTTAGTGGGAGATTTATGATGCTTACGCC 624  
DB 481 GTATGCTGTTCTCTGAGTGAAGTTTTTGTGGGAGCATTTATGAGGCTTACGCC 540  
QY 625 AATTTGGCGTGAAGTTATGATGATGACATTAACCGCTGATTAATACCGGGGTG 684

DB 541 AATTTGAGTGAAGCATCATGATGACGATTAACGTCGCTGATTCATACCGGGTA 600  
QY 685 ATGTCATGACATTAATTAATGCGCATTTGTAAGGCAAGAGCCAAACGACACC 744  
DB 601 ATGTCATGACATTAATTAATGCGCATTTGTAAGGCAAGAGCCAAACGACACC 660  
QY 745 GTTATGCTTTCGCGATGAGAGTCACTATTTCAACCGTTTACGAGG 797  
DB 661 GTTATGCTTTCGCGATGAGAGTCACTATTTCAACCGTTTACGAGG 713

RESULT 4  
US-10-334-143-153/c  
Sequence 153, Application US/10334143  
Publication No. US20040009549A1  
GENERAL INFORMATION:  
APPLICANT: GRIGORIEV, IOR VYACHESLAVOVICH  
TITLE OF INVENTION: METHOD FOR DETECTING REMOTE HOMOLOGUES AND NOVEL  
KINASES IDENTIFIED WITH THE METHOD  
FILE REFERENCE: 038602/1543  
CURRENT APPLICATION NUMBER: US/10/334,143  
CURRENT FILING DATE: 2002-12-31  
PRIOR APPLICATION NUMBER: 60/343,169  
PRIOR FILING DATE: 2001-12-31  
NUMBER OF SEQ ID NOS: 207  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO: 153  
LENGTH: 8202  
TYPE: DNA  
ORGANISM: *Homo sapiens*  
US-10-334-143-153

Query Match 4.8%; Score 38.2; DB 16; Length 8202;  
Best Local Similarity 52.9%; Pred. No. 2.4;  
Matches 82; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 240 AGAAGCTGTCGATGAGATTTTGTATGACAGCAGCAGCATTTGTCGGCAAAATCTCA 299  
DB 5953 AGAAGCTTTCATATGATTTTAAAGCTGTGGAAGTATTTGTTTCAATAGAAAGTCT 5894  
QY 300 TACTTATCCATTTTATCAGAGATCGCGTGTATGAGCAACCCCAACGCGATCGAA 359  
DB 5893 GCGCTTTCAGATGACATGCTCCCGCTTTTGTGTGCTTAAATGATTTGTAC 5834  
QY 360 TGGTGAAGTTCAGATTTTCCATTAACCTTTAT 394  
DB 5833 TTTTTCAGAGACTTTCATTTCTGTTACATTTGAT 5799

RESULT 5  
US-10-142-426-426/c  
Sequence 426, Application US/10142426  
Publication No. US20040048333A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: Deforge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Geriltsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

;; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
;; FILE REFERENCE: P330R1C24  
;; CURRENT APPLICATION NUMBER: US/10/142,426  
;; CURRENT FILING DATE: 2002-05-09  
;; Prior Application removed - See File Wrapper or Palm  
;; NUMBER OF SEQ ID NOS: 550  
;; SEQ ID NO 426  
;; LENGTH: 747  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-10-142-426-426

Query Match 4.4%; Score 35; DB 13; Length 747;  
Best Local Similarity 7.7%; Pred. No. 5.5; Mismatches 349; Indels 0; Gaps 0;  
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

QY 130 TCTATCCAGTTTACTGTGTAACGCTCTAATCCAAATCGCCAGATTGTGTAATGG 189  
DB 554 YMW..K.B.G.DGW.SR.HEKMB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495  
QY 190 GCATTGCTGGCGGAGTAGTGTATGACGAAGATATGACCGCTCATGTGAGAACCTGTC 249  
DB 494 DC.R.D.NS.N.BBAA.RT..NRNG.C.DBAGH.AC.W..CR..NSTYY.....TYG.. 435  
QY 250 GATGAGGATTTTGTATGACGAGCAGCAGTATTGTTCGCAAAAAGTCCATCTTATCT 309  
DB 434 GW.T..HB.W..MENT..N..C.AABDSNBS.MT.T..HR..MWCD..R.MAC.CM.BT 375  
QY 310 AATTATACGCGATCCGCTGTGTATGACGACCCCAAGCGATCGGATGTGGAGT 369  
DB 374 ..S..BR.NSG.H.TSSGH.YCB..DHN..AM.MM..HNSN.NMC.D....WYDDT.Y 315  
QY 370 GTCAGTATTTCCATTCAGCTTTATTAAACCGTGGAATGTCAAAATAGAAATTT 429  
DB 314 H..ANTS...H..ND.B..R..MTH.SN..KCS.Y.YA.SA.GHCMSG.A..WC.CT 255  
QY 430 GGTATCCAGCCCGGAGCGGCTAATGTGTATCTTCACTTACTTCAAGAAGAAATG 489  
DB 254 T.TA..SNB.NCAM..ABASYD.KT.SB.S.MSYHCDMB.TKCSY....S.RSNECMK 195  
QY 490 CCGGTGGCTTTGATCATGTCGCGCAATTCAGCATCGTGCGCAAAATTAACGGCTGG 549  
DB 194 YG.KTHGCBTC.Y..CHB.MT.HYVA.T.HDSDSC.SR.NB.W..CSNBTH..YY. 135  
QY 550 GTGAATACACATCCGCTGACTATTCTTATGAAAAAGATTTTATGCGGATATT 609  
DB 134 A...HHHR.C.HRYRNMWS.R.CC.NH..CM.CRHRMNY.H.HM.B.RTYN.M..HC 75  
QY 610 ATTGATGCTTACGCGCAATTTGGCGTCGAGTATGCGATGCCATTAACGCCGCTTG 669  
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTMSBC..SR.H.HACB..TKRB...G...YM 15  
QY 670 ATCAATAC 677  
DB 14 .HY.RHHM 7

RESULT 6  
US-10-123-155-426/C  
Sequence 426, Application US/10123155  
Publication No. US20030068794A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: Deforge, Laura  
APPLICANT: Deenoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Geritsen, Mary B.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven

;; APPLICANT: Smith, Victoria  
;; APPLICANT: Stewart, Timothy A.  
;; APPLICANT: Tumas, Daniel  
;; APPLICANT: Watanabe, Colin X  
;; APPLICANT: Wood, William  
;; APPLICANT: Zhang, Zemin  
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
;; FILE REFERENCE: P330R1C20  
;; CURRENT APPLICATION NUMBER: US/10/123,155  
;; CURRENT FILING DATE: 2002-04-15  
;; Prior Application removed - See Palm or File Wrapper  
;; NUMBER OF SEQ ID NOS: 550  
;; SEQ ID NO 426  
;; LENGTH: 747  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-10-123-155-426

Query Match 4.4%; Score 35; DB 15; Length 747;  
Best Local Similarity 7.7%; Pred. No. 5.5; Mismatches 349; Indels 0; Gaps 0;  
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

QY 130 TCTATCCAGTTTACTGTGTAACGCTCTAATCCAAATCGCCAGATTGTGTAATGG 189  
DB 554 YMW..K.B.G.DGW.SR.HEKMB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495  
QY 190 GCATTGCTGGCGGAGTAGTGTATGACGAAGATATGACCGCTCATGTGAGAACCTGTC 249  
DB 494 DC.R.D.NS.N.BBAA.RT..NRNG.C.DBAGH.AC.W..CR..NSTYY.....TYG.. 435  
QY 250 GATGAGGATTTTGTATGACGAGCAGCAGTATTGTTCGCAAAAAGTCCATCTTATCT 309  
DB 434 GW.T..HB.W..MENT..N..C.AABDSNBS.MT.T..HR..MWCD..R.MAC.CM.BT 375  
QY 310 AATTATACGCGATCCGCTGTGTATGACGACCCCAAGCGATCGGATGTGGAGT 369  
DB 374 ..S..BR.NSG.H.TSSGH.YCB..DHN..AM.MM..HNSN.NMC.D....WYDDT.Y 315  
QY 370 GTCAGTATTTCCATTCAGCTTTATTAAACCGTGGAATGTCAAAATAGAAATTT 429  
DB 314 H..ANTS...H..ND.B..R..MTH.SN..KCS.Y.YA.SA.GHCMSG.A..WC.CT 255  
QY 430 GGTATCCAGCCCGGAGCGGCTAATGTGTATCTTCACTTACTTCAAGAAGAAATG 489  
DB 254 T.TA..SNB.NCAM..ABASYD.KT.SB.S.MSYHCDMB.TKCSY....S.RSNECMK 195  
QY 490 CCGGTGGCTTTGATCATGTCGCGCAATTCAGCATCGTGCGCAAAATTAACGGCTGG 549  
DB 194 YG.KTHGCBTC.Y..CHB.MT.HYVA.T.HDSDSC.SR.NB.W..CSNBTH..YY. 135  
QY 550 GTGAATACACATCCGCTGACTATTCTTATGAAAAAGATTTTATGCGGATATT 609  
DB 134 A...HHHR.C.HRYRNMWS.R.CC.NH..CM.CRHRMNY.H.HM.B.RTYN.M..HC 75  
QY 610 ATTGATGCTTACGCGCAATTTGGCGTCGAGTATGCGATGCCATTAACGCCGCTTG 669  
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTMSBC..SR.H.HACB..TKRB...G...YM 15  
QY 670 ATCAATAC 677  
DB 14 .HY.RHHM 7

RESULT 7  
US-10-146-731-426/C  
Sequence 426, Application US/10146731  
Publication No. US20030129692A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: Deforge, Laura  
APPLICANT: Deenoyers, Luc

```
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Mei-Qiang
APPLICANT: Gerltisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Thomas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C323
CURRENT APPLICATION NUMBER: US/10/146,731
PRIOR APPLICATION: 2002-05-15
NUMBER OF SEQ ID NOS: 550
LENGTH: 747
TYPE: PRT
ORGANISM: Homo Sapien
US-10-146-731-426
```

```
Query Match 4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;
```

```
QY 130 TCTATCCAGGTTTACTGCTGTAACCGCTTAATCCAAATCGCCGCAATGCTGTAATGG 189
DB 554 YMW..K.B.G.DGM.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
QY 190 GCATTCCTGGGGAGTAGTATGACGAAGATATGACCGCTCAGTGGAGAACCTGTC 249
DB 494 DC.R.D.NS.N.BBAI.RT..NRKNG.C.DBAQH.AC.M..CR..NSTYY.....TYG.. 435
QY 250 GATGAGATTTTGTATGACGAGACGACGATTTGTGCGCAAAAGTCATTAATCTT 309
DB 434 GW.T..HB.W..MNT..N..C.AABDSNBS.MT.T..HR..MWCDD..R.MAC.CM.BT 375
QY 310 AATTTATACGATCCGCTGCTGTTGATGCAACCCCAAGCGATCCGAATGTTGAGT 369
DB 374 S...BR.NSG.H.TSSGH.YCB..DHNH..AM.MM..HNSN.NNC.D...WYDDT.Y 315
QY 370 GTCAGATTTTCCATTAACCGCTTATTAACCGTGAATGTCAACAAATAGAGATTTT 429
DB 314 H..ANTS...H..ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255
QY 430 GGTATGACCCCGAGCGCGCTAATGTTGATCTTCACTTAATCTCAAGAGAAATG 489
DB 254 T.TA..SNB.NCAM..ABASYBD.KT.SB.S.MSYHCDMBB.TKCSY....S.RSNBCKM 195
QY 490 CCGCTGCTTTTATGATGATGCGCAAAATTCGATGCGTGGCAAAATTAAGCGCTGCG 549
DB 194 YG.KTHHCGNBT.C.Y..CHB.MT.HYWA.T.HGDSGC.SR.NB.W..CSNBTH..Y.Y. 135
QY 550 GTTGAATACACATCCGCTGATCTATTTTCTATTAGAAAAGATTTTATGTCGGATAT 609
DB 134 A...HHHR.C.HRYRBNWS.R.CC.NH..CM.CMEHMMY.H.HM.B.RTYN.M..HC 75
QY 610 ATTGATGCTACGCAAAATTTGGCGCGCAAGTTAATGATGACATTAACCGCGCTG 669
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTPMSBC..SR.H.HACB..TKRB...G...YM 15
QY 670 ATCAATAC 677
DB 14 .HY.RHMM 7
```

RESULT 8  
US-10-140-472-426/c

```
Sequence 426, Application US/10140472
Publication No. US20030138868A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Mei-Qiang
APPLICANT: Gerltisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Thomas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C168
CURRENT APPLICATION NUMBER: US/10/140,472
PRIOR APPLICATION: 2002-05-06
NUMBER OF SEQ ID NOS: 550
LENGTH: 747
TYPE: PRT
ORGANISM: Homo Sapien
US-10-140-472-426
```

```
Query Match 4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;
```

```
QY 130 TCTATCCAGGTTTACTGCTGTAACCGCTTAATCCAAATCGCCGCAATGCTGTAATGG 189
DB 554 YMW..K.B.G.DGM.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
QY 190 GCATTCCTGGGGAGTAGTATGACGAAGATATGACCGCTCAGTGGAGAACCTGTC 249
DB 494 DC.R.D.NS.N.BBAI.RT..NRKNG.C.DBAQH.AC.M..CR..NSTYY.....TYG.. 435
QY 250 GATGAGATTTTGTATGACGAGACGACGATTTGTGCGCAAAAGTCATTAATCTT 309
DB 434 GW.T..HB.W..MNT..N..C.AABDSNBS.MT.T..HR..MWCDD..R.MAC.CM.BT 375
QY 310 AATTTATACGATCCGCTGCTGTTGATGCAACCCCAAGCGATCCGAATGTTGAGT 369
DB 374 S...BR.NSG.H.TSSGH.YCB..DHNH..AM.MM..HNSN.NNC.D...WYDDT.Y 315
QY 370 GTCAGATTTTCCATTAACCGCTTATTAACCGTGAATGTCAACAAATAGAGATTTT 429
DB 314 H..ANTS...H..ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255
QY 430 GGTATGACCCCGAGCGCGCTAATGTTGATCTTCACTTAATCTCAAGAGAAATG 489
DB 254 T.TA..SNB.NCAM..ABASYBD.KT.SB.S.MSYHCDMBB.TKCSY....S.RSNBCKM 195
QY 490 CCGCTGCTTTTATGATGATGCGCAAAATTCGATGCGTGGCAAAATTAAGCGCTGCG 549
DB 194 YG.KTHHCGNBT.C.Y..CHB.MT.HYWA.T.HGDSGC.SR.NB.W..CSNBTH..Y.Y. 135
QY 550 GTTGAATACACATCCGCTGATCTATTTTCTATTAGAAAAGATTTTATGTCGGATAT 609
DB 134 A...HHHR.C.HRYRBNWS.R.CC.NH..CM.CMEHMMY.H.HM.B.RTYN.M..HC 75
QY 610 ATTGATGCTACGCAAAATTTGGCGCGCAAGTTAATGATGACATTAACCGCGCTG 669
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTPMSBC..SR.H.HACB..TKRB...G...YM 15
```

QY 670 ATCATAC 677  
DB 14 .HY.RHMM 7

RESULT 9  
US-10-141-761-426/c  
Sequence 426, Application US/10141761  
Publication No. US20030148432A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C198  
CURRENT APPLICATION NUMBER: US/10/141,761  
CURRENT FILING DATE: 2002-05-08  
Prior Application removed - See Palm or File Wrapper  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 426  
LENGTH: 747  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-141-761-426

Query Match 4.4%; Score 35; DB 15; Length 747;  
Best Local Similarity 7.7%; Pred. No. 5.5;  
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

QY 130 TCATCCAGTTTACTGTGAAACGCTTAATCCAAATGCGCAGATTGCTAATG 189  
DB 554 YMW..K.B.G.DGM.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495  
QY 190 GCATGCGTGGCGGATAGTATGACGAAGATATGACCGCTCATGCTGAGAACCTGTC 249  
DB 494 DC.R.D.NS.N.BBA.RT..NRKNG.C.DBAGH.AC.M..CR..NSTY.....TYG.. 435  
QY 250 GATGAGATTGTTGATGACGACGACGATATTGTGCGCAAAAGTCCATCTTACTCT 309  
DB 434 GW.T..HB.W..MNT..N..C.AABDSNBS.MT.T..HR..MDCD..R.MAC.CM.BT 375  
QY 310 AATTATTCAGCGATCCGCTGTTGATGACGACCCCAAGCGATCCGATGTTGAGAT 369  
DB 374 .S...BR.NSG.H.TSSGH.YCB..DHNM..AM.MM...HNSN.NMC.D...WYDPT.Y 315  
QY 370 GTGAGTATTCAGCTTACGCTTATTAACCCGTGGAATGTCAAAAGTAAAGATT 429  
DB 314 H..ANTS....H..ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255  
QY 430 GGTATCCAGCCCGGCGGCTAATGTTGATGATCTTACTTACTCAAGAGAAATG 489  
DB 254 T..TA..SNB.NCAM..ABASYB.KT.SB.S.MSYHCDRMB.TKCSY...S.RSNBCK 195  
QY 490 CCGTGCTTTGATGATCGCGCAATTCAGATGCGGCAAAATTAATGCGCTGCG 549  
DB 194 YG.KYHCGMBTC.Y..CHB.MT.HYVA.T.HDSDSC.SR.NB.W..CSNBTH..Y.Y. 135  
QY 550 GTTGAATACACATCGGTGATCTATTTCATTGAAAAAGTTTATTGCGCGGATATT 609

DB 134 .A...HEHR.C.HRYRBNWS.R.CC.NH..CM.CENHRMRY.H.HM.B.RTN.M..HC 75  
QY 610 ATTATGCTTACGCGCAATTTGGGCTGCAAGTAAATGCAATGACCATTAACCCCTG 669  
DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTMSBC..SR.H.HACB..TKRB...G...YM 15

QY 670 ATCATAC 677  
DB 14 .HY.RHMM 7

RESULT 10  
US-10-142-885-426/c  
Sequence 426, Application US/10142885  
Publication No. US20030157604A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C248  
CURRENT APPLICATION NUMBER: US/10/142,885  
CURRENT FILING DATE: 2002-05-10  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 426  
LENGTH: 747  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-142-885-426

Query Match 4.4%; Score 35; DB 15; Length 747;  
Best Local Similarity 7.7%; Pred. No. 5.5;  
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

QY 130 TCATCCAGTTTACTGTGAAACGCTTAATCCAAATGCGCAGATTGCTAATG 189  
DB 554 YMW..K.B.G.DGM.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495  
QY 190 GCATGCGTGGCGGATAGTATGACGAAGATATGACCGCTCATGCTGAGAACCTGTC 249  
DB 494 DC.R.D.NS.N.BBA.RT..NRKNG.C.DBAGH.AC.M..CR..NSTY.....TYG.. 435  
QY 250 GATGAGATTGTTGATGACGACGACGATATTGTGCGCAAAAGTCCATCTTACTCT 309  
DB 434 GW.T..HB.W..MNT..N..C.AABDSNBS.MT.T..HR..MDCD..R.MAC.CM.BT 375  
QY 310 AATTATTCAGCGATCCGCTGTTGATGACGACCCCAAGCGATCCGATGTTGAGAT 369  
DB 374 .S...BR.NSG.H.TSSGH.YCB..DHNM..AM.MM...HNSN.NMC.D...WYDPT.Y 315  
QY 370 GTGAGTATTCAGCTTACGCTTATTAACCCGTGGAATGTCAAAAGTAAAGATT 429  
DB 314 H..ANTS....H..ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255  
QY 430 GGTATCCAGCCCGGCGGCTAATGTTGATGATCTTACTTACTCAAGAGAAATG 489



```
Db 254 T.TA..SNB.NCAM..ABASYD.KT.SB.S.MSYHCDMB.TKCSY....S.RSBNCKM 195
Qy 490 CCGCTGCTTTGATCATGTCGGCAATTGACGATCGTGGCAAAATTAAGCGCTGG 549
Db 194 YG.KYHHCNBT.C.Y..CHB.MT.HYWA.T.HDGSQC.SR.NB.W..CSNBTH..YY. 135
Qy 550 GTTGAATACACATCCGTGCTACTATTTCATTAAGAAAAGATTTTAAAGCGGATAT 609
Db 134 A...HHNR.C.HRYRWN.S.R.CC.NH..CM.CMHRMY.H.HM.B.RTYN.M..HC 75
Qy 610 ATTGATGCCGACCCAAATTGGCGTCGAAGTTAATCCGATACCATTAAGCGCTTG 669
Db 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSB.C.SR.H.HACB..TKRB...G...YM 15
Qy 670 ATCAATAC 677
Db 14 .HY.RHMM 7

RESULT 11
US-10-158-790-426/c
; Sequence 426, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P333081C448
; CURRENT APPLICATION NUMBER: US/10/158,790
; CURRENT FILING DATE: 2002-05-30
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 426
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-158-790-426

Query Match 4.4%; Score 35; DB 15; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

Qy 130 TCTATCCAGGTTTACTGTGTAAGCGCTTAATCCAAATCGGCCAGATTGGTAATGG 189
Db 554 YWM..K.B.G.DGM.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
Qy 190 GCATTGCTGGCGGATAGTATGACGAAGATATGACCGCTCATGTGTGAGAACTGTC 249
Db 494 DC.R.D.NS.N.BBAH.RT..NRKYG.C.DBAGH.AC.M..CR..NSTYY....TYG.. 435
Qy 250 GATGAGATTTGATGACGACGACGACGATATTTTGTGGCAAAAAGTCATATTATCT 309
Db 434 GW.T..HB.W..MHT..N..C.AABDNBS.MT.T..HR..MMCD..R.MAC.CM.BT 375
Qy 310 AATTATCAGGACCGCTGTGTGATGACCAACCCCAAGCGATCCGATGTTGAGT 369
Db 374 S...BR.NSG.H.TSGH.YCB..DHMM..AM.MM..HNSN.NMC.D....WYDT.Y 315
```

```
Qy 370 GTGAGATTTCCATACGCTTTATTAACCGGGAATGCAACAATAGAAGATTT 429
Db 314 H..ANTS...H..ND.B..R..MTH.SN...KCS.Y.YA.SA.GHCWGG.A..WC.CT 255
Qy 430 GGTATGACCCCGAGCGCGCTTAATGTGATCTTCACTTACTCAAGAAGAATG 489
Db 254 T.TA..SNB.NCAM..ABASYD.KT.SB.S.MSYHCDMB.TKCSY....S.RSBNCKM 195
Qy 490 CCGCTGCTTTGATCATGTCGGCAATTGACGATCGTGGCAAAATTAAGCGCTGG 549
Db 194 YG.KYHHCNBT.C.Y..CHB.MT.HYWA.T.HDGSQC.SR.NB.W..CSNBTH..YY. 135
Qy 550 GTTGAATACACATCCGTGCTACTATTTCATTAAGAAAAGATTTTAAAGCGGATAT 609
Db 134 A...HHNR.C.HRYRWN.S.R.CC.NH..CM.CMHRMY.H.HM.B.RTYN.M..HC 75
Qy 610 ATTGATGCCGACCCAAATTGGCGTCGAAGTTAATCCGATACCATTAAGCGCTTG 669
Db 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSB.C.SR.H.HACB..TKRB...G...YM 15
Qy 670 ATCAATAC 677
Db 14 .HY.RHMM 7

RESULT 12
US-10-137-871-426/c
; Sequence 426, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P333081C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See File Wrapper
; SEQ ID NO 426
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-871-426

Query Match 4.4%; Score 35; DB 16; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

Qy 130 TCTATCCAGGTTTACTGTGTAAGCGCTTAATCCAAATCGGCCAGATTGGTAATGG 189
Db 554 YWM..K.B.G.DGM.SR.HBKWB.NDMY.AB..N.AA.A.SBBS.NK.H.....TSD 495
Qy 190 GCATTGCTGGCGGATAGTATGACGAAGATATGACCGCTCATGTGTGAGAACTGTC 249
Db 494 DC.R.D.NS.N.BBAH.RT..NRKYG.C.DBAGH.AC.M..CR..NSTYY....TYG.. 435
```

Oy 250 GATAGAGATTGGTATGAGGAGAGACAGCATATTGGTCGGCAAAAAGTCCATACCTTATCT 309  
 Db 434 GM.T..HB.W..MHT..N..C.AABDSNB.S.MT.T..HR..MCCD..R..MAC.CM.BT 375  
 Oy 310 AATTATTCAGGCATCCGCTGGTATGAGCAACCCCAACGGATCCGATGTTGGAGT 368  
 Db 374 .S...BR..NSG.H.TSSGH.YCB..DHNM..AM.MM...HNSN.NMC.D...WYDDT.Y 315  
 Oy 370 GTGAGTATTCCTCATACGCTTTATTTATTAACCCGTGGAATGTCAAACTAGAGATTT 429  
 Db 314 H..ANIS...H..ND..B..R..MTH.SN...KCS.Y.YA.SA.GHCMSG.A..WC.CT 255  
 Oy 430 GGTATTCACCCCGGAGCGCGCTAATTGGTTGATCTTCATCTTACTACTCAAGAGAAATG 489  
 Db 254 T.TN..SNB.NCAM..ABASVBD.KT.SB.S.MSVHCRMB.TKRSY...S.RSNBOMK 195  
 Oy 490 CCGCTGGCTTTTGATCATCTGCGGCAAAATTCAGCATCGTGCGCAAAAATTCGCGCTGCG 549  
 Db 194 YG.KYHHCGBTC.Y..CHB.MT..HYMA.T.HDGSQC.SR.NB..W..CSNBTH...YY 135  
 Oy 550 GTTGAAATACATCCGCGGTACTATTATTTATGAAAAAGATGTTTGTGGCGGATATT 609  
 Db 134 .A...HHNR.C.HRPRRMS.R.CC.NH..CM.CMRHRMY.H.HM.B.RTYN.M..HC 75  
 Oy 610 ATTGATCCCTACGCCCAAAATTGGCGTGAAGTTAATGCATGACCATTTAAACGCGCGTTG 668  
 Db 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSBC..SR.H.HACB..TKRB...G...YM 15  
 Oy 670 ATCAATAC 677  
 Db 14 .HY.RHHN 7

```

RESULT 13
US-10-140-923-426/C
; Sequence 426, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tunas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330301C186
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 426
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-923-426

Query Match          4.4%; Score 35; DB 16; Length 747;
Best Local Similarity 7.7%; Pred. No. 5.5;
Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0

130 TCTATCCAGGTTTACTGGTGAACGCTTAATCCAAATCGCCAGATTGTGTAATGG 189

```

```

Db      554 YMW..K.B.G.DGM..SR..HEKXB..NDMY..AB..N.AA.A.SBBS..NK.H.....TSD 435S
Oy      190 GCATTGCGTGGCGGGATAGTGTATGATACGAAGATATAGCCGCTCATGTGTGAGAACTGTG 249
Db      494 DC.R.D.NS.N.BBA..RT..NRKXG.C.DBAGH.AC.M..CR..NSTYY.....TYG.. 435S
Oy      250 GATGAGGATTTTGTATGCAGCAGGAGAGAGATTTGTGTGGCAAAAAGTCATTTACTTATCT 309
Db      434 GW.T..HB.W..MENT..N..C.AABDSNBS..MT.T..HR..MMCD..R..MAC..CM..BT 375
Oy      310 AATTATTATGACGATCCGCTGGTTGATGCAACCCCAACCGATCCGATGCTGGAGT 365S
Db      374 ..S...BR..NSG..H.TSSG..YCB..DHNM..AM.MK...HNSN.NMC.D...WYDDT.Y 315S
Oy      370 GTGAGATTTCCTCATTCAGCTTATTTAAACCCGAGATGTCAACAATATGAAGATTT 422
Db      314 H..ANTS...H..ND.B..R..MTH..SN...KCS.Y.YA..SA.GHCMSG.A..WC..CT 255
Oy      430 GGTATGACACCCGAGCGCGCTAATTGTTGATGTTCTTACTTACTTCAAGAAGAAATG 488S
Db      254 T.TA..SNB.NCAM..ABASVBD..KT.SB.S.MSHCDRMB..TICSY...S.RSNBOMK 195S
Oy      490 CCGCTGGCTTTTGATCATGTGCGCGAAATTAAGCATGCGTGGCGAAAAATTAACGCGCTGG 543S
Db      194 YG..KYHNCGNATC.Y..CHB..MT..HYMA..T..HDGDSGC..SR..NB..W..GSNBTH..YY 135S
Oy      550 GTTGAAADACATCCGTGTGTACTATTTCATTGAAAAAAGTTTATAGTGGCGGATATT 605S
Db      134 ..A...HHNR.C.HRRRWMS..R.CC.NH..CM.CNHRIMMY..H.HM.B..RTYN.M..HC 75
Oy      610 ATTGATGCCATGCCCAAAATTTGGCGTCAAGTTATGCGATGACCATTTAAACGCGCTTG 665S
Db      74 DTN.NN.NNH.D.M.T..MM.T..Y..YSSSTAMSBC..SR.H.HACB..TKBB..G...YM 15
Oy      670 ATGCAATAC 677
Db      14 ..HY.RHHM 7

```

```

, RESULT 14
, Sequence 426, Application US/10141756
, Publication No. US20030207359A1
, GENERAL INFORMATION:
, APPLICANT: Baker, Kevin P.
, APPLICANT: Beresini, Maureen
, APPLICANT: DeForge, Laura
, APPLICANT: Desnoyers, Luc
, APPLICANT: Filvaroff, Ellen
, APPLICANT: Gao, Wei-Qiang
, APPLICANT: Gerritsen, Mary E.
, APPLICANT: Goddard, Audrey
, APPLICANT: Godowski, Paul J.
, APPLICANT: Gurney, Austin L.
, APPLICANT: Sherwood, Steven
, APPLICANT: Smith, Victoria
, APPLICANT: Stewart, Timothy A.
, APPLICANT: Tumas, Daniel
, APPLICANT: Watanabe, Collin K
, APPLICANT: Wood, William
, APPLICANT: Zhang, Zemin
, TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
, TITLE OF INVENTION: ACIDS ENCODING THE SAME
, FILE REFERENCE: P3330R1C200
, CURRENT APPLICATION NUMBER: US/10/141,756
, PRIOR FILING DATE: 2002-05-08
, PRIOR Application removed - See File Wrapper or Palm
, NUMBER OF SEQ ID NOS: 550
, SEQ ID NO 426
, LENGTH: 747
, TYPE: PRT
, ORGANISM: Homo Sapien

```

US-10-141-756-426

Query Match 4.4%; Score 35; DB 16; Length 747;

Best Local Similarity 7.7%; Pred.No.5.5; Prior Application removed - See File Wrapper or Palm

Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

DB 130 TCTATCCAGGTTTACGTGGTGAAGCGCTTAATCCAAATCCGCGAGATGTTGTAATG 189

DB 554 YMW.K.B.G.DGW.SR.HBKWB.NDMY.AB.N.AA.A.SBBS.NK.H.....TSD 495

DB 190 GCATTGCGTGGCGGATAGTGTATGACGAAGATATGACCGCTCATGTTGGAACCTGTC 249

DB 494 DC.R.D.NS.N.BBA.RT.NRKNG.C.DBAGH.AC.M..CR..NSTYV.....TYG.. 435

DB 250 GATGAGATTGATGACGAGACGATATTTGCGCAAAAAGTCCATATCTATCCT 309

DB 434 GW.T..HB.W..MENT..N.C.AABDSNBS.MT.T..HR..MCCDD..R..MAC.CM.BT 375

DB 310 AATTATACGAGATCCGCTGTTGATGACCAACCCCAACGCGATCCGATGTTGAGT 369

DB 374 S...BR.NSG.H.TSSGH.YCB..DHNW..AM.MW...HNSN.NMC.D....WYDPT.Y 315

DB 370 GTGAGATTTCCTATGAGCTTTATTAACCGGTGATGTCACAAATAGAAATTT 429

DB 314 H..ANTS...H..ND.B..R..MTH..SN...KCS.Y.YA.SA.GHCCMSG.A..WC.CT 255

DB 430 GGTATGACCCCGAGCGGCTAATGTTGATCTTCACTTACTTCAAGAAGAATG 489

DB 254 T.TA..SNB.NCAM..ABASVD.KT.SB.S.MSHCDMB.TKCSY...S.RSNBCK 195

DB 490 CCGCTGCTTTGATCATGTGCGCAATTCAGCATGCGTGCAAAAATTAACGCGCTGCG 549

DB 194 YG.KYHCGNBT.C.Y..CHB.MT.HYWA.T.HDGSBC..SR.NB..W..CSNBTH..YY. 135

DB 550 GTTGAATACACATCCGCTGATCTTATTTCAATGAAAAGTTTGTAGCGGATAT 609

DB 134 A...HHR.C.HRRRWS.R.CC.NH..CM.CRHMMY.H.HM.B.RTYN.M..HC 75

DB 610 ATTGATGCTTACGCAAAATTTGGCGTGAAGTTAATGACATTAAGCGCGCTG 669

DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSBC..SR.H.HACB..TKRB...G...YM 15

DB 670 ATCAATAC 677

DB 14 HY.RHEM 7

RESULT 15

US-10-141-756-426/c

Sequence 426, Application US/10141759

Publication No. US20030207361A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Beresini, Maureen

APPLICANT: DeForge, Laura

APPLICANT: Desnoyers, Luc

APPLICANT: Filvaroff, Ellen

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K

APPLICANT: Wood, William

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME

FILE REFERENCE: P3330R1C197

CURRENT APPLICATION NUMBER: US/10/141,759

CURRENT FILING DATE: 2002-05-08

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 426

LENGTH: 747

TYPE: PRT

ORGANISM: Homo sapien

US-10-141-756-426

Query Match 4.4%; Score 35; DB 16; Length 747;

Best Local Similarity 7.7%; Pred.No.5.5; Prior Application removed - See File Wrapper or Palm

Matches 42; Conservative 157; Mismatches 349; Indels 0; Gaps 0;

DB 130 TCTATCCAGGTTTACGTGGTGAAGCGCTTAATCCAAATCCGCGAGATGTTGTAATG 189

DB 554 YMW.K.B.G.DGW.SR.HBKWB.NDMY.AB.N.AA.A.SBBS.NK.H.....TSD 495

DB 190 GCATTGCGTGGCGGATAGTGTATGACGAAGATATGACCGCTCATGTTGGAACCTGTC 249

DB 494 DC.R.D.NS.N.BBA.RT.NRKNG.C.DBAGH.AC.M..CR..NSTYV.....TYG.. 435

DB 250 GATGAGATTGATGACGAGACGATATTTGCGCAAAAAGTCCATATCTATCCT 309

DB 434 GW.T..HB.W..MENT..N.C.AABDSNBS.MT.T..HR..MCCDD..R..MAC.CM.BT 375

DB 310 AATTATACGAGATCCGCTGTTGATGACCAACCCCAACGCGATCCGATGTTGAGT 369

DB 374 S...BR.NSG.H.TSSGH.YCB..DHNW..AM.MW...HNSN.NMC.D....WYDPT.Y 315

DB 370 GTGAGATTTCCTATGAGCTTTATTAACCGGTGATGTCACAAATAGAAATTT 429

DB 314 H..ANTS...H..ND.B..R..MTH..SN...KCS.Y.YA.SA.GHCCMSG.A..WC.CT 255

DB 430 GGTATGACCCCGAGCGGCTAATGTTGATCTTCACTTACTTCAAGAAGAATG 489

DB 254 T.TA..SNB.NCAM..ABASVD.KT.SB.S.MSHCDMB.TKCSY...S.RSNBCK 195

DB 490 CCGCTGCTTTGATCATGTGCGCAATTCAGCATGCGTGCAAAAATTAACGCGCTGCG 549

DB 194 YG.KYHCGNBT.C.Y..CHB.MT.HYWA.T.HDGSBC..SR.NB..W..CSNBTH..YY. 135

DB 550 GTTGAATACACATCCGCTGATCTTATTTCAATGAAAAGTTTGTAGCGGATAT 609

DB 134 A...HHR.C.HRRRWS.R.CC.NH..CM.CRHMMY.H.HM.B.RTYN.M..HC 75

DB 610 ATTGATGCTTACGCAAAATTTGGCGTGAAGTTAATGACATTAAGCGCGCTG 669

DB 74 DTN.NN.NMH.D.M.T.MM.T.Y..YSTAMSBC..SR.H.HACB..TKRB...G...YM 15

DB 670 ATCAATAC 677

DB 14 HY.RHEM 7

Search completed: April 30, 2004, 23:19:44

Job time : 369.368 Secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 08:18:22 ; Search time 2238.64 Seconds

(without alignments)  
10745.552 Million cell updates/sec

Title: US-10-603-260-4

Perfect score: 555

Sequence: 1 atcgagatattcttctgtca.....cttcacagagaatgctgtc 555

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapept 1.0

Searched: 3470272 seqs, 2167151995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vt.\*  
15: em\_da.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_ov.\*  
22: em\_ov.\*  
23: em\_ov.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vt.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_inv.\*  
32: em\_hcg\_other.\*  
33: em\_hcg\_mus.\*  
34: em\_hcg\_pln.\*  
35: em\_hcg\_rtd.\*  
36: em\_hcg\_mam.\*  
37: em\_hcg\_vte.\*  
38: em\_sy.\*  
39: em\_hcgo\_hum.\*  
40: em\_hcgo\_mus.\*  
41: em\_hcgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213.8	38.5	247450	1	AP005349
2	213.8	38.5	302331	1	AE016810
3	191.4	34.5	10977	1	AE004352
4	178.8	32.2	339350	1	AP005088
5	64.8	11.7	622	1	VC0231080
6	37.4	6.7	178879	2	AC124052
7	37.2	6.7	2000	6	AX655393
8	36.8	6.6	26997	6	AX251505
9	36.8	6.6	183397	10	AL592225
10	36.4	6.6	209061	2	AC130676
11	36.2	6.5	54762	2	AC130294
12	36	6.5	101342	8	AP003850
13	35.8	6.5	271370	2	AC112046
14	35.8	6.5	27243	3	CE006547
15	35.8	6.5	215278	10	AC065652
16	35.6	6.4	115218	9	HS15941
17	35.6	6.4	170260	2	AL158204
18	35.6	6.4	212449	2	AL355972
19	35.6	6.4	224330	2	AC113079
20	35.2	6.3	671	8	AY135373
21	35.2	6.3	169572	2	AC023832
22	35	6.3	128769	2	AL139261
23	35	6.3	170110	9	CNS01RHT
24	35	6.3	215747	10	AC117581
25	34.8	6.3	623	3	AY047332
26	34.6	6.2	7218	6	166494
27	34.6	6.2	103513	9	AL357713
28	34.6	6.2	198105	2	AC026160
29	34.6	6.2	233018	2	AC113788
30	34.6	6.2	237754	2	AC117125
31	34.6	6.2	288092	2	AC118793
32	34.4	6.2	2277	6	AR004980
33	34.4	6.2	244169	2	AC107357
34	34.4	6.2	245308	2	AC108307
35	34.4	6.2	262126	2	AC098502
36	34.2	6.2	3596	8	AF096277
37	34	6.1	131027	9	AC004806
38	33.8	6.1	236470	10	AL831793
39	33.8	6.1	51531	2	AC115089
40	33.8	6.1	143589	10	AC124745
41	33.8	6.1	174508	9	AC021561
42	33.8	6.1	177781	9	AC114689
43	33.8	6.1	268699	2	AC139758
44	33.4	6.0	1629	8	FSY420192
45	33.4	6.0	55096	2	AC100894

# ALIGNMENTS

RESULT 1  
LOCUS AP005349 247450 bp DNA linear BCT 04-DEC-2003  
DEFINITION Vibrio vulnificus Y016 DNA, chromosome II, complete genome,  
section 6/8.  
ACCESSION AP005349 BA000038  
VERSION AP005349.1 GI:37201329  
KEYWORDS  
SOURCE  
ORGANISM  
Vibrio vulnificus Y016  
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
Vibrionaceae; Vibrio.  
REFERENCE  
1 Chen, C.Y., Wu, K.M., Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L.,  
Lin, Y.M., Chen, H.J., Shen, A.B., Li, D.C., Su, T.D., Shao, C.P.,

TITLE Lee, C.T., Hor, L.I. and Tsai, S.F.  
JOURNAL Comparative Genome Analysis of *Vibrio vulnificus*, a Marine Pathogen  
REFERENCE Genome Res. 13, 2577-2587 (2003)  
AUTHORS 2 (bases 1 to 247450)  
JOURNAL Chen, C.Y., Wu, K.M. and Tsai, S.F.  
TITLE Direct Submission  
JOURNAL Submitted (29-May-2002) Shih-Feng Tsai, National Health Research  
Institute, Division of Molecular and Genomic Medicine, 128  
Yen-Chu-Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China  
(E-mail: pctsai@nhri.org.tw, Tel: 886-2-8146-1041,  
Fax: 886-2-2789-0484)  
COMMENT This sequence was determined by the Sequencing Core of the National  
Yang-Ming University Genome Research Center (YMGRC;  
http://genome.ym.edu.tw)

FEATURES  
source  
location/Qualifiers  
1..247450  
/organism="Vibrio vulnificus Y016"  
/mol\_type="genomic DNA"  
/strain="Y016"  
/db\_xref="taxon:196600"  
/chromosome="11"  
complement(73..1092)  
/gene="VVA1126"  
complement(73..1092)  
/gene="VVA1126"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein N, non-specific porin"  
/protein\_id="BAC97152.1"  
/db\_xref="GI:37201330"  
/translation="MNKNITLALVASATFGTQAVAVEVYNDGTTFSIGHVSVALD  
SKQDLGVAVSPRINFNATIDLGNGFPADAKGWSINYLDGSEALSTRIGYGLTH  
ESLGGVGTQAWPYSVAGVADMPAFANDPFIYODHNLGTGRAEEMVSYANFELG  
BAGKALIGLGGGKRTDNNNSYGNRAOIALNYSIANYTANYATGVDNNGVGGTGT  
SSSHITSATGSGYGGIYACGYANNEHNSGDCMLDTSAYEALAAATYTNININS  
VYVESVIDEELSKRYSTSAVOAETNFTPKVGFAGYQFDLQSGSDTKERADKRWLG  
ARYYL"  
complement(1331..1519)  
/gene="VVA1127"  
complement(1331..1519)  
/gene="VVA1127"  
/note="identified by Glimmer2"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC97153.1"  
/db\_xref="GI:37201331"  
/translation="MRALYVMDISLKLASRGYLGIRPAFNPFSIKKGESISVCNII  
INISVFLITPNVIGVCP"  
1559..2050  
/gene="VVA1128"  
1559..2050  
/gene="VVA1128"  
/codon\_start=1  
/transl\_table=11  
/product="thiol-disulfide isomerase and thioredoxins"  
/protein\_id="BAC97154.1"  
/db\_xref="GI:37201332"  
/translation="MKIVILKEMLPVLAVALVISALDFYSSNLPKRNALPLSAMT  
TLGTQVLDLSAQSGSPVVYFWATCGPCKVSPPIINADSPYISVAMSGDEGSV  
NOYLOSQGYDVTINDPTGKISRDNQGVTPVIVIRNGKIEFTISGVSPIGLMRL  
WLS"  
2203..3180  
/gene="VVA1129"  
2203..3180  
/gene="VVA1129"  
/codon\_start=1  
/transl\_table=11  
/product="putative outer membrane protein"  
/protein\_id="BAC97155.1"  
/db\_xref="GI:37201333"  
/translation="MKKALATPAIISAVVTGSSPFAATYTKTDGTELUKIGAVFERGCP  
IGTDGVEIDGTMEDSTRARLNKGTTLGNDLQAFGYEABQKTGESSFNRRVYAG

gene  
CDS  
/gene="VVA1130"  
3606..5381  
/note="identified by GeneMark and Glimmer2"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC97156.1"  
/db\_xref="GI:37201334"  
/translation="MYDGRSKTVANSKYLLIKAYIVAVLITQTLVDHETLTPSY  
VFAIMVALVSPSHKIKVICALIATSTLFGGVDPIDVIDIEFTILFLPCYLL  
PFGTIAVFAVALLTLASYFTVPPEGLDELIDALELVISAFAIMVYQKFIYQL  
TKYREDSRDYTLQLENRAFYSDINQVSLAQQRNCLNIDIDAFKEINSPGHE  
QADLILKRSRLRKRTLGFGQVYEMGEDEPAFLVYKESINDEVALDILKRALTA  
APFLERGSYQISVATGIALFNSMTNLDIWLNSDIAQAKATRGSGVARYDEALD  
ETIRHQIEREDKALTELKQSLVQPKVITITGKISITGTGAELIRMOHPELITTPPE  
ITIAEKTKOITIGRWVTEAACQAKKYLIDGKIPVSNVSTQVPAHDTILOVETA  
LASHLPASITQDEITETVLKGPCKLIEITCNLCBOGTVAIIDPGIAYSSLNTRKO  
LPIDVLKIDKAFIDDCVDNHDHMLVRIIVQLGHNMCKVLAEGVETEEQRLLVSBQ  
CDEFGYLYSKPLDVEEFNRDLQI"  
complement(5356..5475)  
/gene="VVA1131"  
complement(5356..5475)  
/gene="VVA1131"  
/note="identified by Glimmer2"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC97157.1"  
/db\_xref="GI:37201335"  
/translation="MSIPKPHNHSKGNHAKESWCKKMTQGRKAVYTG"  
5718..6596  
/gene="VVA1132"  
5718..6596  
/gene="VVA1132"  
/codon\_start=1  
/transl\_table=11  
/product="putative transporter, DME family"  
/protein\_id="BAC97158.1"  
/db\_xref="GI:37201336"  
/translation="MQSCHPILIGASWMLTAGIAFAIVNSLAQVASTIOPAIESTVAL  
IQYGLIVILPYLTKTIGIRQSLRTSHFPMHALRVFLSYIGQLMVALAVVPLWQ  
IALMTSPFATIGSLGLKKEKVGARWCAATLFGFGAMIIDEPADDFHALTLFVQ  
AAPFAVCYSLMKKLSVNDPSTMTVVYLLITTPNIMATIDPMTSTIWLILFA  
AGCMTLAQMALYKAYASDASFVQPFDAKIPINVLAGWLVGWPSPRMLWGAFTI  
IASVARTMETKTK"  
6661..7266  
/gene="VVA1133"  
6661..7266  
/gene="VVA1133"  
/codon\_start=1  
/transl\_table=11  
/product="glutaredoxin-related protein"  
/protein\_id="BAC97159.1"  
/db\_xref="GI:37201337"  
/translation="MSEAIKTIPLYMAGSMGPKVNIPOGECTLRKIDITDFANELN  
GIPELEVDOMLSHMEPLKGLGMHAPILIVGKVISQGEALNRGLISVIREWTKR  
DILKNIIVFGKATCPYCVAKOLDSAGIHGYHDVVKESALRYMIPVKAIIGKTI  
PTVVPQIWLDSGYITGGCDLLEYLIKERGLDVIPNVVEAN"  
complement(7360..7494)  
/gene="VVA1134"  
complement(7360..7494)  
/gene="VVA1134"  
/note="identified by Glimmer2"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC97160.1"

```

/db_xref="GI:37201338"
/translation="MGVELGAGYIGQMAMKAMKMYKATKAKKAFISMKMKAKK"
gene
  complement(7650, .8000)
  /gene="VVA1135"
  complement(7650, .8000)
  /gene="VVA1135"
  /note="identified by GeneMark and Glimmer2"
  /codon_start=1
  /transl_table=11
  /product="conserved hypothetical protein"
  /protein_id="BAC97161.1"
  /db_xref="GI:37201339"
  /translation="MQONEPQVYKQIQOENLPKALELKACDDEEVAEATSLSQ
  FALAEVNGEKIHYHTLOENAEAGETEVEHVHNGEHLVXFAAMFVDVFEIKAKET
  YQAGKTYQPKRS"
gene
  complement(8059, .9252)
  /gene="VVA1136"
  complement(8059, .9252)
  /gene="VVA1136"
  /note="major facilitator superfamily"
  /codon_start=1
  /transl_table=11
  /product="permease"
  /protein_id="BAC97162.1"
  /db_xref="GI:37201340"
  /translation="WKSLOKPLMLMIIATGQGVATYLPSPILSHDLGTQAD
  VQNVVTFVGFAGSOLFVGBIPDIAIGRPITLIGITVIGVITICAFSSIEALEL
  GRLOQAGASVYSGSYRPSYRGNQIKALSYITSLIPPIHVPVGGATVH
  SQVAFVFTCYLALFTLTGTVLPETLPKRRFRPLVATYLRITATPQYTGAS
  FNMNYSVLSVLSVFPFMQKGFSAADYGSWMLPSAGLLGSTVAVVHLRVEY
  KRLAASFVIALSGGLIVAFSSVLIATFMSPAQMSPLSILALHKNQNA
  GSVALSQSVQMVAAALLGFLVHCVRTOQLFYILVAGCFAVTLKQVMTPE
  EVQAV"
gene
  9360, 10247
  /gene="VVA1137"
  9360, 10247
  /gene="VVA1137"
  /codon_start=1
  /transl_table=11
  /product="transcriptional regulator"

Query Match      38.5%; Score 213.8; DB 1; Length 247450;
Best Local Similarity 64.1%; Pred. No. 2.2e-57;
Matches 339; Conservative 0; Mismatches 187; Indels 3; Gaps 1;

```

```

QY 421 ATACCTTAAAGGCGCGACCGCTTTACTGCGGATCCACTG--GACTCCAGATGATACC 477
DB 146149 ATTCACATGGGTGACAGACCCCTTCTTACATGATGATACACACAAATGCTATGGGACG 146208
QY 478 GTCAAGATAGAGAAACCAAAATAGCTGGCATGACAGATCATATGCCGA 526
DB 146209 GTCCAGGTGAGAGACAAAGATCTCAGTATGACTGACCAATGTTTGA 146257

RESULT 2
AE016810
LOCUS
DEFINITION
  Vibrio vulnificus CMCP6 chromosome II section 3 of 6 of the
  complete sequence.
ACCESSION
  AE016810 AE016796
VERSION
  AE016810.1 GI:27358548
KEYWORDS
  'Vibrio vulnificus CMCP6
  Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
  Vibrionaceae; Vibrio.
  1 (bases 1 to 302331)
  Kim, Y.R., Lee, S.E., Kim, C.M., Kim, S.Y., Shin, E.K., Shin, D.H.,
  and Rhee, J.H., Choy, H.E., Progliske-Fox, A., Hillman, J.D., Handfield, M.
  Characterization and Pathogenic Significance of Vibrio vulnificus
  Antigens Preferentially Expressed in Septicemic Patients
  Infect. Immun. 71 (10), 5461-5471 (2003)
  14500463
  2 (bases 1 to 302331)
  Jeong, H., Moon, Y.H., and Kim, J.J.
  Direct Submision
  Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong,
  Yuseong-gu, Daejeon 305-811, South Korea
  3 (bases 1 to 302331)
  Rhee, J.H., Kim, S.Y., Chung, S.S., Lee, S.E. and Choy, H.E.
  Direct Submision
  Submitted (13-DEC-2002) Department of Microbiology, Genome Research
  Center for Enteropathogenic Bacteria, Chonnam National University
  Medical School, Hak-1-Dong, Dong-gu, Kwang-Ju 501-746, South Korea
  1.302331
  location/Qualifiers
  /organism="Vibrio vulnificus CMCP6"
  /mol_type="genomic DNA"
  /strain="CMCP6"
  /db_xref="taxon:216895"
  /chromosome="II"
  complement(70, .1368)
  /locus_tag="VV20549"
  complement(70, .1368)
  /locus_tag="VV20549"
  /note="COG3325"
  /codon_start=1
  /transl_table=11
  /product="chitinase"
  /protein_id="AA007498.1"
  /db_xref="GI:27358549"
  /translation="MKKTLITPAALFSAFAGSAVNOEKVAVGFADMDQYANSPYTVK
  DIPAEKLTHTVIAFLSMCGPHTGASETVQKVAACQCEKEPFSALIVDTAEALKEKDFG
  KVSVPKRGHIDQALAEIKOHPDLKILPSFGWMTSPFPAMADPYAMQFSTKAT
  ELIKOYDFPDGIDLMVEYGGGGLTSPWNPATLSDBQKAERDAFYLVKTRADL
  DALEKTRREVELSTRAVGVGAKAADIPIKAAOPYLTMMFAMTYVEIIGMGQOTGHTTN
  LHATRSWNGKADVFINQTELGIPSKRLVYGAATFYRGCMQGTQDTYGOAPKEDILS
  EQGAPFGTGENGEYFMFLVKNYGEKQGYVNDYDQSAPLYMNEKRVFISFEDQRS
  IKAKQMAQKSLNGIFTWELSGDPSGRLVDJMAEMTSK"
  complement(1510, .2502)
  /locus_tag="VV20550"
  complement(1510, .2502)
  /locus_tag="VV20550"
  /note="contains sigma factor-related N-terminal domain;
  COG3390"
  /codon_start=1

```





Db 259437 ATGAAATCATTTATATACAGGTTATATATAGCATGGTTAGTCATGCAACTCTCAGC 259496  
QY 61 CATCGTCGCATTAATTTGGGTTATGCTAAACCATTTAGTACACTGACTGCTATC 120  
Db 259497 TTGGCGCTTAAAAAATTTGGGCTATCAAAACAAAGCTCCTACACACGCTCAATTT 259556  
QY 121 GATGATGAGCCATTTTTCGCCCTTGACCGATGCTCATCTCATGCTGCTTAATGCT 180  
Db 259557 GATGAGAAAAAGTTTCCATTCATTAAGCAAAAGGCTCAGACAGACGCGGAAATGTG 259616  
QY 181 TTATGTCGACACAGTTTGGGCGGATTTGGTATCAAAAGTTATCTTAAATTCGGCGCACCG 240  
Db 259617 TTGGTTGGCATAGTTTGGGAGGCTTGAATGATCAAAATTTCTTCCCTCAGTCCGCT 259676  
QY 241 TCTGTGAAACCTCTCCCATGTGTCGCGCATGCGCTCACTTTCAGAGAGCTTCAT 300  
Db 259677 ACCCTTGACCAATCTCAATGTGTGCGCATAGGCTCACCGCTCAAAAGCGCTTCCATC 259736  
QY 301 GTCAATTAATTAATGACATTTAGGTTAAGGGGTGCACTAGTAAATTCAGCAAAATTTGGG 360  
Db 259737 GTTACAAAATTCAGAGGCTTGTGTTAAGGTGCGATGTGGCAATTCGCAAAACATGGG 259796  
QY 361 TTAAGAAACAGACAGACGATCCCGCTATCCAAATTCAGGAGATTTGCGAGAACG 420  
Db 259797 CTCATATGCAATGAGAGATGTTGGATTTTCTCTCAAAAGCTTGGCATGATTTGCGAGC 259856  
QY 421 ATACCTTTTATGAGGCTGCGACGCTTTTACTGCGAGTCCAGTCTG---GACTCCGATGTTAC 477  
Db 259857 ATTCATTTGGGTGCAAGACCTCTTCTTCTGATGATACACAGACAAATGCTGATGGAGCG 259916  
QY 478 GTCACTATTAAGAAACCAAAATTAAGTGCATGACAGATCATATGGCA 526  
Db 259917 GTTACAGGTGAGAGACAAAGATATGAGTATGACTGACATGTTTGA 259965

RESULT 3  
AE004352/c 10977 bp DNA linear BCT 10-OCT-2003  
LOCUS Vibrio cholerae O1 biovar eltor str. N16961 chromosome II, section 9 of 93 of the complete chromosome.  
ACCESSION AE004352 AE003853  
VERSION AE004352.1 GI:9657475  
KEYWORDS  
SOURCE  
ORGANISM  
Vibrio cholerae O1 biovar eltor str. N16961  
Vibrio cholerae O1 biovar eltor str. N16961  
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.  
REFERENCE  
AUTHORS 1 (bases 1 to 10977)  
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Selinger, P., McDonald, L., Uitterback, T., Fleischmann, R.D., Snelman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.  
DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae  
NATURE 406 (6795), 477-483 (2000)  
JOURNAL  
MEDLINE 20406833  
PUBMED 10952301  
REFERENCE  
AUTHORS 2 (bases 1 to 10977)  
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Selinger, P., McDonald, L., Uitterback, T., Fleischmann, R.D., Snelman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.  
Direct Submission  
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA  
FEATURES  
source  
location/Qualifiers  
1..10977  
/organism="Vibrio cholerae O1 biovar eltor str. N16961"

gene  
CDS  
/mol\_type="genomic DNA"  
/strain="N16961"  
/serotype="O1"  
/db\_xref="taxon:243277"  
/chromosome="II"  
/note="biotype: E1 Tor"  
/gene="VCA0090"  
complement (81..974)  
/gene="VCA0090"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF96004.1"  
/db\_xref="GI:9657476"  
/translation="MSLVNLLQLCLAIWGSFLFMRIRANSFPAATYIEARVPA  
LSLFLVAQLRLSLRHPHPLILGLINTVPLFAVVALTINVTLSLSTAP  
IMGAVTGLFMTGTPLSRKAVALIGSVAVIVGDVAIGHHALPMVCAALAA  
YGLATVYTKQAPOLSAFENAGSMWACLMVAPLMFVFLRETPSLLEMGAVILGVI  
CTGALVLYRVLKATGAATSLVTPFLFVFGIMVGYLIDEPILGINTLFGTLVLVAG  
TMLVTFPSLRNRLSRQPH"  
1204..1500  
/gene="VCA0091"  
1204..1500  
/gene="VCA0091"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF96005.1"  
/db\_xref="GI:9657477"  
/translation="MARPKIPRIEHCPPASCFKGVPIROLARVELADEAURL  
VDQLGLQOOALQMOVSRQTLANLVKARFVVDCLHOKALYIQADINKSSD"  
complement (1671..2300)  
/gene="VCA0092"  
complement (1671..2300)  
/gene="VCA0092"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF96006.1"  
/db\_xref="GI:9657478"  
/translation="KIVTLHGLVHGLVMOPLAQLNKLGQTEVISTNTLAIDEX  
VQITDSALQORINVLVGHSLGIMIKYLSRHSFNVIVGVVALPLKASIVP  
KIQQLGLAMLGNAHYGLIHDQSMELRQDGLAGTLRFSPRLILGSGQCDGT  
VTAETQISMTDHLHGHSTGLVYVSHKTAQIDYFIRHNGFHKKIP"  
2400..2573  
/gene="VCA0093"  
2400..2573  
/gene="VCA0093"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF96007.1"  
/db\_xref="GI:9657479"  
/translation="MEQPCWLSVNPITWFKYVNEDETCLLPVPRLAKLEKXH  
DDCLFKRSQDDM"  
2801..3157  
/gene="VCA0094"  
2801..3157  
/gene="VCA0094"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF96008.1"  
/db\_xref="GI:9657480"  
/translation="MDVNIKSPRIPIINDVLAQLERTSTKVTIHLHLSSESARF  
GELYSLNGISAKMLSHRKMLEIGWVREDDPTTPVTVYIGLTKSGRVYANGLYRM  
IDHVVENEREPLESN"

gene 3426. 4124  
/gene="VCA0095"  
CDS 3426. 4124  
/gene="VCA0095"  
/note="identified by Glimmer2; putative"  
/codon start=1  
/transl table=11  
/product="hypothetical protein"  
/protein\_id="AAF96009.1"  
/db\_xref="GI:9657481"  
/translation="MYQKLSHMLFLSVFLVFLGILSLTYTSTPTPTNLPDRQEH  
VAPKPAEIEIHEHPAVELESTEENRVNVEQEVQSGEAPRVIEPQNLQVWLP  
NTPPELLADLKPSQAVTDIOALEFENIDQFRTVAGDSATLPLNGESIRYLIEACQ  
OAGQIQTAGNFETQGGTFPVTFTGQSHIMGFIHQSGIKIEGVHQAQWVQVPS  
HSEHDSHDBAH"  
4167. 5390  
/gene="VCA0096"  
CDS 4167. 5390  
/gene="VCA0096"  
/note="identified by Glimmer2; putative"  
/codon start=1  
/transl table=11  
/product="hypothetical protein"  
/protein\_id="AAF96010.1"  
/db\_xref="GI:9657482"  
/translation="MKKSLALAVSVGLCSPAIMATTEIDVLYPTDAGFRQEHV  
AOWHSHVANKVILKOSGLDILKYNLAATKREVOYDPOGLKKSQSEVLDAEPNRID  
APADVAAYQGVADWVAIFRYLDVNNSPDERQNGSYISGLGLAVASAVKPYQ  
NAKSVSHSYLNEGCAETFEIHEGNHGNHAYQVELEPHNNNGTVDVAGVIGK  
QRTITAYPHILFVGVSRYKSSPNACBAPGCVKQYANSTRALIGLIPHAQYTG  
KPVVDGNGFDTEPDNTNNVFTIKGPLAPDMKILTLPIVASAQSSSTAVALDITH  
EYRGDLSIFLAPDGSYVWLKQANRYDRQSYNVQFTLNDVPSAAEGEMLEIDHR  
GGLKGLNQFQITFP"  
/gene="VCA0097"  
CDS complement(5403. 6119)  
/gene="VCA0097"  
complement(5403. 6119)  
/note="identified by Glimmer2; putative"  
/codon start=1  
/transl table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF96011.1"  
/db\_xref="GI:9657483"  
/translation="MTVTIDMTCLRLAPKQIEFTLLVKSNNRPPDCGMAIPGQWFD  
EULSAQSGPADKPSASARRICQKIHYPNFIISDPVDANPKRDSGWSISLHYA  
LHNSVNTQIEBAENDKSRVNFALETILQGVAFVHVOIQHAWKQKRAAYEYS  
VVLFSLEFLVADIIIAVAKFQVDVNRMTVRRLLIDTGVIVSTKIAASSKGGKXP  
ATVYRLASNEVYFQCLAKG"  
5518. 7825  
/gene="VCA0098"  
CDS 5518. 7825  
/gene="VCA0098"  
/note="similar to GB:U05568 SP:PI8133 PID:147307 GB:U00096  
PID:1651456; identified by sequence similarity; putative"  
/codon start=1  
/transl table=11  
/product="nicotinate phosphoribosyltransferase"  
/protein\_id="AAF96012.1"  
/db\_xref="GI:9657484"  
/translation="MNPRLFSPHIIIRSLDDAYKINMOAIIHFPYDVSVYELIAR  
SESDAGLDAROBIAHGLTFRSDADHVTQAPHLKATFLQSLREYFHPQEOY  
ENGIYQGGKQOLRISIRGSRDITLVELVVAIVSEVRSQRMAYEVADLPKLYKT  
KIDQKAEIERGINNFSITGEMTRRSQVQVLAICQOEIPQWVLGTSNHFAR  
EDIKPRTGTAHEMWHQALYENRDSQOVALERPLAFDMGLAIPDTLITDAFIN  
DNRRLANAIDVGRDSCGPFKGMKTAHYQQLIDPTTLFTISDGLDDQALETC  
EYFAGKVKISFGITFLINDLANWRKPAQGVETRPISITKLAECQGRVAKLSQPER  
AMCEDEPIFLANKREFNIELVDVALIQELRHQKRSRPHYISA"  
7856. 8896  
/gene="VCA0099"  
CDS 7856. 8896  
/gene="VCA0099"  
/note="similar to GB:U00096 PID:1742678 PID:1742686  
PID:1742694 PID:1742718; identified by sequence

similarity; putative"  
/codon start=1  
/transl table=11  
/product="oxidoreductase, Gfc/Idh/MocA family"  
/protein\_id="AAF96013.1"  
/db\_xref="GI:9657485"  
/translation="MNPPLKVGIVIGYSAKTFHLPIIDTLESLTSAISSQOQV  
OEDMPOIAPDSACRLITSDVDVITLPPNNVHPFIADLAEHGIVVEKPRTOI  
EQRTELEBARQGLISVFNHRMDGDELVTKLEQGLQDVKTFESHPDRFVY  
RQWRQADQGGSLFDLAPHLDQALVFGIPQSLSDCKMRKRPATLVDYDLOIY  
YPOHVVRLHANTLYPSPNVRYOVLGSLGKXVGLDPOBRLKAGRPHPQWSQMP  
EQQGILYHAGNENVTTEIGYQHYFTQVVEALRNQASNPVSMSALOSIOLIELALE  
SSAKGQRLAVTL"  
/gene="VCA0100"  
CDS complement(8933. 9304)  
/gene="VCA0100"  
complement(8933. 9304)  
/note="similar to GP:7226016; identified by sequence

Query Match 34.5%; Score 191.4; DB 1; Length 10977;  
Best Local Similarity 60.5%; Pred. No. 2,6e-50;  
Matches 315; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

1 ATGCAATTATCTTGTTCATGCACTCTATATGCAATGCTGGTAAATGCAATCCGCTTACT 60  
2300 ATGAAAGATTGTATATACGACGCGTTTGATATGCAATGCTTGTATGCAACCTTCGCG 2241  
61 CATGCTGTGATTAATTGGGTTATGCAATCAACCAATTAGTACAACTCACTGCTATC 120  
2240 CAACGTCATATTAACCTGGATATACAGACGAAAGTGATCAGCTTCAACACTTATGCGCATC 2181  
121 GATGATGAGCCATTTTCCGCCGCTTGAACGATGCTCACTGATGCTTCCGCTTAATGCT 180  
2180 GATGATGAAGATATTTGACATGACATGACAGGCACTAGCAACAAATGCAATCAACGTCG 2121  
181 TTATGCGACACAGTTTGGCGGATGCGATGATGAACGTTAATGCAATGCGCGACG 240  
2120 TTGGTAGACATGATGTTGGTGGATGATGATGAACATTAACCTTACGTTTCTCAACCT 2061  
241 TCCGTGAAACCCCTCCCATGTCGTCGCGCATGCGCTCACTTGGACAGAGCTTCATT 300  
2060 TCACCCAAATGATTTCTCATGATGTTGCTTGGCTCCCTCCCTTAAAGGGCATGATT 2001  
301 GTCATTAATTAATGCAATTAATGTTTGGGATGCGGCTGACTAGTAAATCAGAGAAATTTGGG 360  
2000 GTCCCTTAAGTCCAGACGTTTGGGCTAGGAGCTATGTTAGTAAACCACTTCTGATGGC 1941  
361 TTAAGAAACACGACGACGATCCCGCTATCCACAAATTCAGGACGATATTGACGAGACG 420  
1940 TTGCAATGATCAAGATGATGCTGGAAATGGCCACAAAGGCTGGTTGCAATGACAGAAACA 1881  
421 ATACCTTTAGGCTGGCGACGCTTTTACTGCGGATCCACTGATCCGATGCTACCGTC 480  
1880 TTGGGTTTCGAAATTTGGCCCAATCTTACCTCGGAGGCTCAGGAATGTGTATGACAGCGTC 1821  
481 ACACTAGAGAAACCAAAATAGCTGCGCATGACGATCATAT 521  
1820 ACGGTTGCCAGACTCAAAATCTCAGGAATGACGATCATCT 1780

RESULT 4  
AP005088 339350 bp DNA linear BCT 05-MAR-2003  
LOCUS Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 5/6.  
DEFINITION AP005088 BAO00032  
ACCESSION AP005088.1 GI:28809504  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Vibrio parahaemolyticus  
Vibrio parahaemolyticus  
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
Vibrionaceae; Vibrio.  
REFERENCE  
1 Nesu,H., Iida,T., Sugahara,T., Yamachi,Y., Park,K.S., Yokoyama,K.,

TITLE Makino, K., Shingawa, H. and Honda, T.  
A filamentous phage associated with recent pandemic Vibrio  
parahaemolyticus O3:K6 strains  
JOURNAL J. Clin. Microbiol. 38 (6), 2156-2161 (2000)  
MEDLINE 20295086  
PUBMED 10834969  
REFERENCE 2

AUTHORS Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,  
Ragomori, K., Iijima, Y., Najima, M., Nakano, M., Yamashita, A.,  
Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shingawa, H.,  
Hattori, M. and Iida, T.  
TITLE Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
distinct from that of V. cholerae  
JOURNAL Lancet 361 (9359), 743-749 (2003)  
MEDLINE 22508454  
PUBMED 12620739

TITLE 3 (bases 1 to 339350)  
Oshima, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T.,  
Honda, T., Shingawa, H., Hattori, M. and Iida, T.  
JOURNAL Direct Submission  
Submitted (09-ARR-2002) Ken Kurokawa, Osaka University, Genome  
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,  
Japan (E-mail: ken-gen-info.osaka-u.ac.jp/  
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,  
Fax: 81-6-6879-2047)  
COMMENT This clone was isolated from a patient presenting with acute  
gastroenteritis.

FEATURES  
source Location/Qualifiers  
1..339350  
/organism="Vibrio parahaemolyticus"  
/mol\_type="genomic DNA"  
/strain="O3:K6"  
/sub\_strain="RIMD 2210633"  
/db\_xref="taxon:670"  
/chromosome="2"  
64..1428  
/gene="VP1137"  
64..1428  
/gene="VP1137"  
/note="similar to GB:AA19314.1 (AE008712) percent  
identity 76 in 437 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="cytochrome BD2, subunit I"  
/protein\_id="BAC62480.1"  
/db\_xref="GI:28809505"  
/translation="MFGIDAPMLARIQPAFTVSPHTITFGATYLAVLEGMLK  
TRNDYQLYHFMKIFAVNMGVSGLVAAVQFGTNGSFDPASTIGPLITVEV  
LTAFFLEAGPLGVMLFGMRVGEKLFHFPATSMVALGTIISTFWILASNMWQPOYE  
IVDRVVPDWFPAIVPNSFVYRLAHMSVAAPSAALFVGSAAWHLRLNOSTAVKT  
MFMSLGIIVLPAIVGVIGVGLNTEHPAKIALIEGHMNSDCKPPLILFGKP  
NMBORNDYALEIEPVLGSLIRHSITRIPALCKPEERDPSPIVMSGRFVVGGL  
LMTQSPYSMTLRKNTLYTSRMFLKSLIPMGSGALAIAGPFTTGVGQPPVYVVG  
QKTRDAVSAHSDLOMSISLCLFVVISLVFGFYTIHQIKGPDALHDESDMTTV  
SGRI"  
1457..2464  
/gene="VP1138"  
1457..2464  
/gene="VP1138"  
/note="similar to GB:AA19315.1 (AE008712) percent  
identity 67 in 336 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="cytochrome BD2, subunit II"  
/protein\_id="BAC62481.1"  
/db\_xref="GI:28809506"  
/translation="MHPDLSTVFPATVPAITMTINDGPDIGILMPITKQKQD  
VMVNSVAPWDGNTWIVLGSALFSGAPPAVAVITATITPILMLAIFGVAFS  
FRFQLEHLKFMRSPIVGLSLTFPFGIVAGVIGFTVENRVFVGSQDLWLVAFS  
IFCGIGVATVALIGSTWLMKTEGALONTMYRPTNTLMLASALIVSAMPIAYP

gene 2469..2627  
/gene="VP1139"  
2469..2627  
/gene="VP1139"  
/codon\_start=1  
/note="unknown"  
/evidence=not\_experimental  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC62482.1"  
/db\_xref="GI:28809507"  
/translation="MIRSEYENQKLVPPVGLDGDWLDGQGCASSGVFINAVSRHDCR  
VILTRHP"  
complement (2684..3439)  
/gene="VP1140"  
complement (2684..3439)  
/gene="VP1140"  
/note="similar to GB:AA96301.1 (AE004375) percent  
identity 62 in 250 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="BAC62483.1"  
/db\_xref="GI:28809508"  
/translation="MKAQIDVFCDELAQGNPCGVBLNWLSDSELLQTHQVAP  
VTSFVKSQSDHIRMFTLAGELNCGSGSGAGALLSKHONVILQSSGYITVS  
ELIDGMVRIYMPGMAKPAHPIPLGISGFNRPIDARTDILVVLSEDDVVCYCNPD  
ELIKIDEFAVIVTAQSSDSGVLYRFPAPKGISDILATGSAQCSLARYMEKLGKKEK  
LTAQLSRSGGVFEVARGSESTIVLSAQVSRV"  
complement (3534..3911)  
/gene="VP1141"  
complement (3534..3911)  
/gene="VP1141"  
/note="similar to GB:AA884740.1 (AE000810) percent  
identity 47 in 116 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="gamma-carboxymuconolactone decarboxylase"  
/protein\_id="BAC62484.1"  
/db\_xref="GI:28809509"  
/translation="MENTRFQTEGELSKIDGAGQVIESINDICPDIAKFTIEYP  
GDIYTRBGDLKSRRIATVSALTAMGNCPTPOLKVALHAAALNVGTBEETKEVILQMSV  
YAGPPAALNGMAFKEVLSEROV"  
complement (3970..4326)  
/gene="VP1142"  
complement (3970..4326)  
/gene="VP1142"  
/note="similar to GB:AAK76923.1 (AE001438) percent  
identity 48 in 109 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="putative transcriptional regulator, MerR family"  
/protein\_id="BAC62485.1"  
/db\_xref="GI:28809510"  
/translation="NMKKEPSSAVGLSYYTLRYEKIGLLKRVHNSSGHRYTTRDI  
DMVNFIKLXKMGMLERIEQYASLRSGSQTIDROKTLSEHNRNLLIHRNQNEHL  
KLEERKINLYKSGKR"  
4561..5874  
/gene="VP1143"  
4561..5874  
/gene="VP1143"  
/note="similar to EMB:CB72498.1 (AL139074) percent  
identity 45 in 406 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="putative molybdenum containing oxidoreductase"

```

/ (protein_id="BAC62486.1"
/db_xref="GI:28809511"
/translation="MSENEKKPKQGIYKYENNPSPKADKVFGRVSPDRSGELKAG
LATMAALGGFIPFRNMPAGIIPAFABGDDVIEKDLTYLNRPMNAEPPLH
LNDVTPRRRHRIRNNNGIIPDNDVNETLITIDGVKPMISTADIKNNDVVEQL
VTECGGNRAFPDPRKSGNWTYGVASSWTGYRLADVKAAYKYGATITAYGD
KHLSEKGLPISRGVPAKAWGSENLIAFQNGEALHPNGAFRLVYRWPSCG
KMLTRIQRDQIHDPKMTGTSYRVNRPVAPGVNAKEDFELIERMVSLLTSPQ
NTEVNGEALIRHAMSGDRKVTQVQISIDGATWMDADLAPANDGAMQTFNAKVF
POAGYEWAKXTDQGVOPFALIMNKYLNNTFRIALIVRS"
5918..6580
/gene="VP1144"
5918..6580
/gene="VP1144"
/ (gene="VP1144"
/ (note="similar to GB:AAK24897.1 (AE005958) percent
identity 39 in 109 aa"
/codon_start=1
/evidence=not experimental
/translation="hypothetical protein"
/product="hypothetical protein"
/protein_id="BAC62487.1"
/db_xref="GI:28809512"
/translation="MRTQPSERVPLAFTLLALSNPVLASNDLNAEKKQAWCK
ACHQEPNGVTVGVPLWGLABRNIASFEGEYSGIDRQHKDAEKLDFLSAPND
FAQGNMVPGVTPDGAARAIATWATNPTPPNMNMTSGLEYSPEGDGLTPGENM
EYVAVCSACHSLHMYTQGLSRQKMDSLDMITEQMEDLSGDDRAILLEYISTYY
GG"
complement(6763..7179)
/gene="VP1145"
complement(6763..7179)
/gene="VP1145"
/ (note="similar to GB:AAQ07974.1 (AE004872) percent
identity 48 in 93 aa"
/codon_start=1
/evidence=not experimental
/translation="hypothetical protein"
/product="hypothetical protein"
/protein_id="BAC62488.1"
/db_xref="GI:28809513"
/translation="MKIKSKNVFVSLFTLLGLGASTKDNVAKGFPVQYAGFPD
GCHSGTKAGSLPDEFKQDISRFNSDKKTAQMSGFRQCSSEQALIRQRTITBEQ
KLIEQRKANEHSRYLLEKHALGVDVTSLESTPK"
complement(7370..7972)
gene
Query Match 32.2%; Score 178.8; DB 1; Length 339350;
Best Local Similarity 60.1%; Pred. No. 5.6e-46;
Matches 316; Conservative 0; Mismatches 207; Indels 3; Gaps 1;
gene
1 ATGCAGATTATCTTGTTCATGAGACTCTATATGATGCTGTTGTAATGATCGCTTAACT 60
120862 ATGAAATCATCATCTTACATGCTTCTTATATGATGAGGCTGTCATGAGCCTTTAGC 120921
61 CATCGTCTGATTAATTTGGGTTATGCTACTCAACACATAGTACAACTCACTGGCTATC 120
120922 CAAGAGCTGGAAAGCTCGGATACGAAGAGCAAGTCTAGCTATACACTTGTCTATC 120981
121 GATGATGAGGCCATTTTTCGCCGCTTGAACCATGCTCACTACATGCTCGCTTAATGCT 180
120982 GATGAGCTTCACTATTTGATTTCTATGACATTCGCTTAATTCACATGACCGCAACG 121041
181 TTAGTGGGACACGTTTGGGCGGATGGGATCAACGTTATCTGAATTCGGGCGGACACG 240
121042 CTGTTGGGACACGCTTGGGAGCTTGATGATTAAGCGCTATCTGCCCAACAAAACCA 121101
241 TCCTGTGAACCCCTCTCCCATGTCGTCGATCGGCTCACTTTGCAAGAGACTTCAT 300
121102 ACCACACAGCTTATCTCTATATGTCGCTATCGGTTGACCTTTAAAGCGCGCTCA 121161
301 GTCAATTAATTTAGCATTTAGGTTTAAAGGATGACATGATTAATTCAGCAATTTGGG 360
121162 GTAGGTCGATTAAGATCTCGATTGGGATGATTTTGAAGAACTCAACACATCAACGA 121221
361 TTTAAAGAACACGACGACGATCCGCTATCCAAATAATGACGATGATTCAGAAACG 420

```

```

Db 121222 TTTAAACATCATCATGATCGTGGGCTTTTCCGCAAAAACCTGGGAGTTCACAGAA 121281
Qy 421 ATACCTTTAGGGGTGGGACGCTTTTACTGCGGATCACTGAGAC---TCCGATGTAC 477
Db 121282 GTCCGATTTGTGCCGCCGCCATTTGTTGATCCGTAAGACAAACCATGTGCGACGTACG 121341
Qy 478 GTACAGTGAAGAAACCAATATGCTGCATGACATCATATCG 523
Db 121342 GTTACGTTAAGAGACAAAGACTGACGCTATGCAAGATCATATCG 121387

RESULT 5
VCH231080 622 bp DNA linear BCT 22-JUL-1998
LOCUS Vibrio cholerae 210r gene.
DEFINITION AJ231080
ACCESSION AJ231080.1 GI:3341555
VERSION
KEYWORDS
SOURCE
ORGANISM
Vibrio cholerae
Vibrio cholerae
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE
1 Fallarino A.
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 622)
TITLE Direct Submission
AUTHORS Submitted (21-JUL-1998) Fallarino A., Dept. of Microbiology and
JOURNAL Immunology, University of Adelaide, SA 5005, Adelaide, AUSTRALIA
FEATURES
source
1..622
/organism="Vibrio cholerae"
/mol_type="genomic DNA"
/strain="Z1/561"
/db_xref="taxon:666"
26..322
/gene="z10r"
26..322
/gene="z10r"
/codon_start=1
/translation="protein"
/protein_id="CA13122.1"
/db_xref="GI:3341556"
/db_xref="SPTREMBL:O87026"
/translation="MLGNALHLYGLQHDQSMELPQRLGCIAGTARFGRPIILGSGM
CDGTVAERQISGMTDHLHLHSHRGVYSHKTAQIDVYFIRHNGFQFKTPE"
ORIGIN
Query Match 11.7%; Score 64.8; DB 1; Length 622;
Best Local Similarity 56.6%; Pred. No. 2.4e-09;
Matches 120; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
gene
310 ATTGAGCAATTAGGTTTAAAGGATGAGCACTAGGTAATTCGACAAATTTGGGTTAAAGAA 369
2 ATCCAGAGTTAGGGCTAGAGCTATGTTAGTAAAGCAATCTGTATGCTTGCAATTG 61
Qy 370 CACGACAGCAATCCCGCTATCCACAAATAACGAGATTTGACAGAACGATACCTTTA 429
Db 62 CATCAATATGCTGGGAATGGCCAAACGCGCTGGTTCATTCAGAGAACATTCGCTTTC 121
Qy 430 GGGCTGGGACGCTTTTATGCGCGATCCACGAGATCGGATGTTACCGTCAAGTAGAA 489
Db 122 GGATTTGCCCAATCTACGCTGAGTCCGAGATGTGATGACACGCTTCC 181
Qy 490 GAAACCAAAATAGCTGGCATGACATCATAT 521
Db 182 GAGACTCAAAATCTCAGGAATGACATCATCT 213
RESULT 6
AC124052/c

```

LOCUS AC124052 178879 bp DNA linear HTG 27-MAR-2003  
 DEFINITION Mus musculus clone RP24-440120, WORKING DRAFT SEQUENCE, 38  
 unnumbered pieces.  
 AC124052.3 GI:29294269  
 VERSION HTG; HTGS PHASE1; HTGS DRAFT.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 178879)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP24-440120  
 Unpublished  
 2 (bases 1 to 178879)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barron, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
 Bouckgeater, B., Brown, A., Camarata, U., Campopiano, A., Chang, J.,  
 Charato, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Fero, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,  
 Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,  
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,  
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunhhang, P., Pierre, N.,  
 Pollard, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,  
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
 Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S.,  
 Theodores, J., Topham, K., Travers, M., Travis, N., Triggilio, J.,  
 Vasiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,  
 Young, G., Zainoun, D., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 178879)  
 Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N.,  
 Anderson, S., Arachchi, H.M., Barron, N., Bastien, V., Bloom, T.,  
 Boguslavsky, L., Bouckgeater, B., Camarata, U., Chang, J., Choepel, Y.,  
 Collamore, A., Cooke, P., Cooke, P., Cornu, B., Dearellano, K.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,  
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,  
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunhhang, P., Pierre, N.,  
 Rachupka, A., Ramasamy, J., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Testaye, S., Theodores, J., Topham, K., Travers, M.,  
 Vasiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, D., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 27, 2003 this sequence version replaced gi:28630062.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WTR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center project name: L19412  
 Center clone name: 440\_120  
 ----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 166751 bases at least Q40  
 Consensus quality: 171820 bases at least Q30  
 Consensus quality: 173712 bases at least Q20  
 Insert size: 148000; agarose-1p  
 Insert size: 175179; sum-of-contigs  
 Quality coverage: 5.6 in Q20 bases; agarose-1p  
 Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 38 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 417 416: contig of 416 bp in length  
 517 516: gap of 100 bp  
 1177 1177: contig of 661 bp in length  
 1178 1277: gap of 100 bp  
 1278 1966: contig of 689 bp in length  
 1967 2066: gap of 100 bp  
 2067 2713: contig of 647 bp in length  
 2714 2813: gap of 100 bp  
 2813 3516: contig of 703 bp in length  
 3517 3616: gap of 100 bp  
 3617 4331: contig of 715 bp in length  
 4332 4431: gap of 100 bp  
 4431 5089: contig of 658 bp in length  
 4432 5189: gap of 100 bp  
 5190 5804: contig of 615 bp in length  
 5805 5904: gap of 100 bp  
 5904 6568: contig of 664 bp in length  
 6569 6668: gap of 100 bp  
 6669 7349: contig of 681 bp in length  
 7350 7449: gap of 100 bp  
 7449 7732: contig of 283 bp in length  
 7733 7832: gap of 100 bp  
 7833 8537: contig of 705 bp in length  
 8538 8637: gap of 100 bp  
 8638 9349: contig of 712 bp in length  
 9350 9449: gap of 100 bp  
 9449 10033: contig of 584 bp in length  
 10034 10133: gap of 100 bp  
 10133 10830: contig of 697 bp in length  
 10831 10930: gap of 100 bp  
 10931 11889: contig of 955 bp in length  
 11890 11989: gap of 100 bp  
 11990 12712: contig of 723 bp in length  
 12713 12812: gap of 100 bp  
 12813 13901: contig of 1089 bp in length  
 13902 14001: gap of 100 bp  
 14001 14946: contig of 945 bp in length  
 14947 15046: gap of 100 bp  
 15047 15701: contig of 655 bp in length  
 15702 15801: gap of 100 bp  
 15802 16508: contig of 707 bp in length  
 16509 16608: gap of 100 bp  
 16609 17266: contig of 658 bp in length  
 17267 17366: gap of 100 bp  
 17367 18500: contig of 1134 bp in length  
 18501 18600: gap of 100 bp  
 18601 19674: contig of 1074 bp in length  
 19675 19774: gap of 100 bp  
 19775 20813: contig of 1039 bp in length  
 20814 20913: gap of 100 bp

```
misc_feature 1. .416 /clone lib=RPci-24 Male Mouse BAC
misc_feature 517. .1177 /note="assembly_fragment"
misc_feature 1278. .1966 /note="assembly_fragment"
misc_feature 2067. .2713 /note="assembly_fragment"
misc_feature 2814. .3516 /note="assembly_fragment"
misc_feature 3517. .4331 /note="assembly_fragment"
misc_feature 4432. .5089 /note="assembly_fragment"
misc_feature 5190. .5804 /note="assembly_fragment"
misc_feature 5905. .6568 /note="assembly_fragment"
misc_feature 6669. .7349 /note="assembly_fragment"
misc_feature 7450. .7732 /note="assembly_fragment"
```

Oy	409	ATTTCAGGAACGATACCTTTAAGGGCTCGGACGCTTTATCTGCGCATCCACGTGACATCC	468
Db	17737	ATCCCGGGCTCGTCCCCCATGGGCTGGGACGCTGCTGAGGGGCTGTGCGCATGCTCTCC	17678
Oy	469	GATGCTACCCGTCACAGTTGGAAGAAACCAAAATAGCTGCATGACATATATTCGGGATA	528
Db	17677	GATGGCACCGTGGGCGCTGGATGAACGCAAGTTCGCCGGGCTGGCCACACACTGCGGATC	17618
Oy	529	TTCCACCA	535
Db	17617	CCGGCCA	17611

Query Match 6.7%; Score 37.2; DB 6; Length 2000;  
Best Local Similarity 7.7%; Pred. No. 2.7;  
Matches 39; Conservative 231; Mismatches 234; Indels 0; Gaps 0.

Oy	30	TATGATGAGCTGGTAAATGATCGCTAGTCAATCGCTGATATAATGGGTATTCGAC	89
Db	523	TMRCAKRRKSYSAARARACMYRKRGGYVAGMMKKRYKRMVYMMKKRYKSKCSWYC	582
Oy	90	TCAAACCATTAAGCTACAACCTACCTCGCTATCGATGATGAGGCCATTTTTCGCCGCTTGA	149
Db	583	KMSYYAACMSARACGAKYCKRCSRSMSSKMSRRCRCKASRYSSAKRYAMMGWTSG	642
Oy	150	CCGATCGCTCACTCATGCTCGCTCGCTTAATGCTTTAGTCGACACAGTTTGGCGGATTTGGT	209
Db	643	SRMSRMSYTCYMRKMSKMSKSTCTMWTYMSRYTAKGSYVRYYRPAACMMWMMWYYRY	702
Oy	210	GATCAAAACGTTATCTAGATCGCGCGACCGCTCTGTGAAACCTCTCCCATGTCGTGC	269
Db	703	RSYNTYAAWYSSTRMAMTKRYSGRWTSYRYKCKSKSWKRSWYVMSWMAATWKKMR	762
Oy	270	CATGGGCTCACTTTGACAAGAGCTTCATGTCAATAAATTAGCATAGGTTTAGG	329
Db	763	RYATRRMMMYRYSMKYTWCTMMGMYWWTYMTMRKRYMYKCTKYVWYSATYWGTM	822
Oy	330	GGTGGCACTAGTAATTACAGCAAAATTTGGTTAAAGAACACAGACGATCCCGCTA	389
Db	823	AAWMAATKORMGMTGAKTRBARAKABYMWMAATWCATKRMWTKGKXMYMTWAKAKMR	882
Oy	390	TCCACAAAATAACAGCAGTATTTGAGAAGATCCATCTTAGGGCTGGCAGACCTTTTACT	449
Db	883	KTYSMRMAATYYTKTRTKYKTCWKKAKRQSWALYMRMMWKGSAKMMWMMGKMGWTKY	942
Oy	450	GCGGGATCCACTGAGCTCCGATGGTACCGTCAACAGTAAAGAAACAAATAAGCTGGCAT	509
Db	943	WYWCCTTWKACGBATKXKMCAGWMAWYSWTRTYWLRTRWRMMWASRTAKRYAPRMMK	1002
Oy	510	GACGAGATCATTCGGCATATCCAC	533
Db	1003	TRAWSKSYARAYWZAGACCTAC	1026

RESULT	8		
AX251505/c			
LOCUS	AX251505	26997 bp	DNA
DEFINITION	Sequence	473 from Patent WO0168912.	linear
ACCESSION	AX251505		
VERSION	AX251505.1	GI:15984928	
KEYWORDS			
SOURCE			
	synthetic construct		

ORGANISM synthetic construct  
artificial sequences.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Epidemiol. Infect. 133: 1-11 (2005)  
Patent: WO 0168912-A 473 20-SEP-2001,  
Epidemiol. Infect. 133: 1-11 (2005)

FEATURES  
source  
1..26997  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 6.6%; Score 36.8; DB 6; Length 26997;  
Best Local Similarity 47.1%; Pred. No. 5.6;  
Matches 113; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 277 TCACCTTGCAGAGGCTTCATTTGCAATAAATTGAGCAATTAGGGTGGCA 336  
DB 13416 TCTCCTGGAACCTACCACTATTAACCTAATAACCTAATAACCTAATTC 13357  
QY 337 CTAGGTAATTCAGCAAGATTGGGTTAAAGACAGACAGCAATCCCTATCCAA 396  
DB 13366 CTACCAACATTTATCACTTCAGCACTAATACCACTCGAACAATCACTTTAAACAA 13297  
QY 397 AATATGAGCAGATTGCGAGAACCAATACCTTAGGGCTGGCCAGCTTTACTGGCGAT 456  
DB 13296 AATAAATTAATAACCTAATAACCTAATAACCTAATAACCTAATAACCTTAC 13237  
QY 457 CCAGTGCAGCTCCGATGCTACCTGCAAGTAGAAGAACCAATAATAGTGGATGACAGAT 516  
DB 13236 CCACCAACACCAATTAACCTCCCACTATTAACCAACCAACCAATTAACCAAT 13177

RESULT 9  
AL592225  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AL592225 183397 bp DNA linear ROD 10-JUN-2002  
Mouse DNA sequence from clone RP23-278M14 on chromosome 1, complete  
sequence.  
AL592225.17 GI:21711841  
HTG.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 183397)  
Blakey, S.  
Direct Submission  
Submitted (10-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail: request@sanger.ac.uk  
humquerry@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk  
On Jul 8, 2002 this sequence version replaced gi:21615572.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP  
database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-278M14 is  
from the RP23-23 Mouse PAC library  
constructed by the group of Peter de Jong.  
For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6.

FEATURES  
source  
1..183397  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/clone="RP23-278M14"  
/clone\_1ib="RP23-23"

ORIGIN

Query Match 6.6%; Score 36.8; DB 10; Length 183397;  
Best Local Similarity 56.7%; Pred. No. 7.6;  
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 288 AGGAGCTTCATTTGCAATAAATTGAGCAATTAGGGTGGCACTAGTAATTC 347  
DB 58507 AGCAGCTTAATGAGCATTAATTAACAGACGAGTGGATGTTCTTCTGTAGCAT 58566  
QY 348 AGCAGATTTGGGTTAAAGAACAGACAGCAATCCGCTATCCACAAATTCAGCAG 407  
DB 58567 GGCTTAATTCAGCTTAAGAACACAGAGGAGAGTCTCTGTAGGCAAACTCAGGTG 58626

RESULT 10  
AC130676/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC130676 209061 bp DNA linear HTG 10-MAR-2003  
Mus musculus clone RP23-167L15, WORKING DRAFT SEQUENCE, 13  
unordered pieces.  
AC130676.3 GI:28894622  
HTG: HTGS PHASE1, HTGS DRAFT.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 209061)  
Birren, B., Nussbaum, C., and Lander, E.  
Mus musculus, clone RP23-167L15  
Unpublished  
2 (bases 1 to 209061)  
Birren, B., Nussbaum, C., Lander, E., All, A., Allen, N., Anderson, S.,  
Barnes, N., Bassett, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Cammarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, J., Diaz, J., S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galasano, J.,  
Gargana, S., Gord, S., Graham, L., Grand-Pierre, N., Hago, B.,  
Horizon, L., Hulme, M., Iliev, I., Johnson, R., Jones, C., Kamet, A.,  
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,  
McCarthy, M., Meldrum, J., Meneses, L., Mihova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Plerre, N., Raymond, C., Retta, R., Rice, C., Rogov, P.,  
Roman, J., Roy, A., Schauer, S., Schuppback, R., Seaman, S., Severy, P.,  
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
Teste, S., Theodore, J., Topham, K., Traversa, M., Vasiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zemlek, L., Zimmer, A. and Zody, M.  
Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 209061)



**AUTHORS**  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Bana, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouknight, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gargyala, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karas, A., Kelis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melgrim, J., Meneus, L., Mihova, T., Mienna, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhag, P., Pierre, N., Rachupka, A., Ramsamy, U., Raymond, C., Rettig, R., Rise, C., Rogov, P., Roman, U., Schauer, S., Schuppach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

**TITLE**  
 Submitted (10-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 10, 2003 this sequence version replaced gi:28269519.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/XM/RepeatMasker.html

**JOURNAL**  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu

**COMMENT**  
 Project information  
 Center project name: 167.L.15  
 Center clone name: 167.L.15  
 Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 205696 bases at least Q40  
 Consensus quality: 205796 bases at least Q30  
 Consensus quality: 207355 bases at least Q20  
 Insert size: 21000; agarose-ftp  
 Insert size: 207861; sum-of-contents  
 Quality coverage: 8.4 in Q20 bases; sum-of-contents  
 Quality coverage: 8.5 in Q20 bases; sum-of-contents

**NOTE:** This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

1 20644: contig of 20644 bp in length  
 20645 20744: gap of 100 bp  
 20745 21744: contig of 1000 bp in length  
 21745 21844: gap of 100 bp  
 21845 23426: contig of 1582 bp in length  
 23427 23526: gap of 100 bp  
 23527 25761: contig of 2235 bp in length  
 25762 25861: gap of 100 bp  
 25862 27654: contig of 1793 bp in length  
 27655 27754: gap of 100 bp  
 27755 33828: contig of 6074 bp in length  
 33829 33928: gap of 100 bp  
 33929 40078: contig of 6150 bp in length  
 40079 49074: gap of 100 bp  
 49075 89252: contig of 49074 bp in length  
 89253 89353: gap of 100 bp  
 89354 110083: contig of 20731 bp in length  
 110084 110183: gap of 100 bp

**FEATURES**  
 source  
 \* 110184 133812: contig of 23629 bp in length  
 \* 133813 133912: gap of 100 bp  
 \* 133813 159643: contig of 25731 bp in length  
 \* 159644 159743: gap of 100 bp  
 \* 159744 205188: contig of 45445 bp in length  
 \* 205189 205288: gap of 100 bp  
 \* 205289 209061: contig of 3773 bp in length.  
 Location/Qualifiers  
 1..209061  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-167L15"  
 /clone\_1b="RP23-167L15"  
 1..20644  
 /note="assembly\_fragment"  
 clone\_end:826  
 vector\_side:left  
 20745..21744  
 /note="assembly\_fragment"  
 /note="assembly\_fragment"  
 23527..25761  
 /note="assembly\_fragment"  
 /note="assembly\_fragment"  
 25862..27654  
 /note="assembly\_fragment"  
 27755..33828  
 /note="assembly\_fragment"  
 33929..40078  
 /note="assembly\_fragment"  
 /note="assembly\_fragment"  
 40179..89252  
 /note="assembly\_fragment"  
 89353..110083  
 /note="assembly\_fragment"  
 110184..133812  
 /note="assembly\_fragment"  
 133913..159643  
 /note="assembly\_fragment"  
 159744..205188  
 /note="assembly\_fragment"  
 205289..209061  
 /note="assembly\_fragment"  
 clone\_end:17  
 vector\_side:right

**ORIGIN**  
 Query Match 6.6%; Score 36.4; DB 2; Length 209061;  
 Best Local Similarity 50.6%; Pred. No. 10;  
 Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Gy 193 AGTTGGCGGATGGATGATCAAGCTTATCTGAATCGCGGACCGTCTGTGAACC 252  
 Db 137780 ATTATAGCGGAAGATGATATGATCTTAAAGCAAAAGCTATGCTGTGAGA 137721  
 253 CTCGCCATGCTGTCGCATCGGCTCACTTTCAGAGAGCTTCAATGTCAATAAAT 312  
 Db 137720 CAGTTCACCTCTTGTCTTATCTTACATTTCCAGAGAACCTTCATATCAAAAT 137661  
 Gy 313 GAGCAATTAAGTTAGGGGCGGACATAGTAATTCAGCAAGAAATTGGGTTAAA 366  
 Db 137660 CAGTCATCTGACCTAGACTTGAACCTGAGTATGTTTCATGCAAGAGATCAGA 137607

**RESULT 11**  
 AC130294 54762 bp DNA linear HTG 19-SEP-2002  
 AC130294  
 LOCUS Homo sapiens chromosome 11 clone RP13-57287 map 11, LOW-PASS  
 DEFINITION  
 SEQUENCE SAMPLING.  
 AC130294  
 AC130294.3 GI:23196586  
 VERSION  
 HTG; HTGS PHASE0.  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 54762)  
Britten, B., Nussbaum, C. and Lander, E.  
Homo sapiens chromosome 11, clone RP13-572P7  
Unpublished  
2 (bases 1 to 54762)

## TITLE

## JOURNAL

REFERENCE  
AUTHORS

Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 54762)

Britten, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barra, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukigalter, B.,  
Camara, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
Cook, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
Horizon, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A.,  
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,  
McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Miwaga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,  
Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J.,  
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

Submitted (19-SEP-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 19, 2002 this sequence version replaced gt:22532570.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: 572\_F\_7

\* NOTE: This record contains 68 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will

\* be preserved.  
1  
706 705: contig of 705 bp in length  
806 805: gap of 100 bp  
1512 1512: contig of 707 bp in length  
1612 1612: gap of 100 bp  
2329 2329: contig of 717 bp in length  
2429 2429: gap of 100 bp  
2430 2430: contig of 719 bp in length  
3149 3149: gap of 100 bp  
3249 3249: contig of 706 bp in length  
3955 3955: gap of 100 bp  
4054 4054: contig of 699 bp in length  
4753 4753: gap of 100 bp  
4853 4853: contig of 710 bp in length  
4854 4854: gap of 100 bp  
5564 5564: contig of 703 bp in length  
5664 5664: gap of 100 bp  
6373 6373: contig of 709 bp in length  
6472 6472: gap of 100 bp  
7181 7181: contig of 709 bp in length  
7281 7281: gap of 100 bp  
7974 7974: contig of 693 bp in length  
8074 8074: gap of 100 bp  
8793 8793: contig of 715 bp in length  
8794 8794: gap of 100 bp  
8893 8893: contig of 708 bp in length  
9601 9601: gap of 100 bp  
9702 9702: contig of 686 bp in length  
10387 10387: gap of 100 bp  
10488 10488: contig of 704 bp in length  
11191 11191: gap of 100 bp  
11281 11281: contig of 716 bp in length  
12007 12007: gap of 100 bp  
12107 12107: contig of 692 bp in length  
12799 12799: gap of 100 bp  
12800 12800: contig of 695 bp in length  
13594 13594: gap of 100 bp  
13694 13694: contig of 724 bp in length  
1418 1418: gap of 100 bp  
14518 14518: contig of 719 bp in length  
15237 15237: gap of 100 bp  
15337 15337: contig of 711 bp in length  
16048 16048: gap of 100 bp  
16149 16149: contig of 705 bp in length  
16853 16853: gap of 100 bp  
16953 16953: contig of 680 bp in length  
17633 17633: gap of 100 bp  
17733 17733: contig of 699 bp in length  
18432 18432: gap of 100 bp  
18533 18533: contig of 712 bp in length  
19244 19244: gap of 100 bp  
19344 19344: contig of 716 bp in length  
20060 20060: gap of 100 bp  
20160 20160: contig of 707 bp in length  
20867 20867: gap of 100 bp  
20967 20967: contig of 705 bp in length  
21672 21672: gap of 100 bp  
21772 21772: contig of 717 bp in length  
22589 22589: gap of 100 bp  
22590 22590: contig of 712 bp in length  
23401 23401: gap of 100 bp  
23415 23415: contig of 714 bp in length  
24215 24215: gap of 100 bp  
24922 24922: contig of 707 bp in length  
25022 25022: gap of 100 bp  
25724 25724: contig of 702 bp in length  
25824 25824: gap of 100 bp  
2619 2619: contig of 695 bp in length  
26519 26519: gap of 100 bp  
27326 27326: contig of 707 bp in length  
27426 27426: gap of 100 bp  
28149 28149: contig of 723 bp in length  
28249 28249: gap of 100 bp  
28957 28957: contig of 708 bp in length  
29057 29057: gap of 100 bp

Query	Best Local Match	Similarity	Score	DB 2	Length
Db	26121	ATGAAGCAGATGATGATATAGACATGGAAGAAAGGAAAGATTGGTTTAA	26180		
Qy	365	AAGAACAAGCAGACGATCCGGTATCCAGAAATCAGGACAGTATTTGACGAACGATAC	424		
Db	26181	AAAAGGGGAGGGGGGAGAAACCATCTCTGTAACAGCCCTTAGATGCTCGAATCACTGCT	26240		
Qy	425	CTTTAGGAGCTGCGAGCCCTTTTACTGCGCATCCATCGATCGATGGTACCGTCA	482		
Db	26241	CTGTAGCCAGGAGGCTGAGTCACTCAGAGATCCATCGCCCTGCTGCTCGC	26298		
RESULT 12	AP003850/c	101342 bp	DNA	linear	PLN 23-UTL-2003
LOCUS	AP003850				
DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7, BAC clone:OJ1793_E11.				
ACCESSION	AP003850				
VERSION	AP003850.2	GI:22415817			
KEYWORDS	Oryza sativa (japonica cultivar-group)				

ORGANISM  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC  
clone:OU1793\_E11  
Published Only in Database (2001)  
2 (bases 1 to 101342)  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Direct Submission  
Submitted (03-UTL-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kamondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaka@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
On Aug 21, 2002 this sequence version replaced g1:14595195.  
Genes were predicted from the integrated results of the following:  
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH  
(http://www.softberry.com/), Genemark hmms  
(http://opal.biology.gatech.edu/Genemark/), GlimmerX  
(http://www.tigr.org/tdb/glimmerx/glmr\_forw.html), RiceHMM  
(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor  
(http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), slm4  
(http://globin.cse.psu.edu/html/docs/slm4.html), gap2  
(http://www.tigr.org/software/glimmerx/), BLASTN and BLASTX. The  
genomic sequence was searched against NCBI Nonredundant Protein  
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA  
sequence database at RGP or DBJ. Protein homologies of the coding  
regions were searched against NCBI Nonredundant Protein database  
with BLASTP. ESTs represent the identified cDNA sequences using  
BLASTN with the corresponding DBJ accession no. and RGP clone ID.  
Full-length cDNAs represent the identified cDNA sequences using  
BLASTN with the corresponding DBJ accession no.

A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with full-length cDNA or  
EST homology (covering almost the entire length of partial  
sequence) is classified as an 'unknown' protein. A gene predicted  
by two or more gene prediction programs is classified as a  
'hypothetical' protein according to IRSP standard. A gene  
predicted by a single gene prediction program is also classified as a  
miscellaneous feature of the sequence.

The orientation of the sequence is from -21M13 to M13rev of the BAC  
clone. This sequence of OU1793\_E11 clone has an overlap with  
OSJNB80036M16 (DBJ: AP005103) clone at 5' end and with OU1657\_A07  
(DBJ: AP003303) at 3' end. The sequence was generated by combining  
Monsanto and RGP-Japan sequencing data. Detailed information on  
overlap and assembly quality together with annotation of this entry  
is available at  
http://rgp.dna.affrc.go.jp/GenomeSeq.html.

Location/Qualifiers

1..101342

source  
organism="Oryza sativa (japonica cultivar-group)"  
mol\_type="genomic DNA"  
cultivar="Nipponbare"  
db\_xref="taxon:39947"  
chromosome="7"  
clone="OU1793\_E11"  
complement(1..1396)  
/gene="OU1793\_E11.101"  
complement(1..1396)  
/gene="OU1793\_E11.101"  
note="3' LTR"  
complement(3384..8953)  
/gene="OU1793\_E11.101"  
complement(3384..8953)  
/gene="OU1793\_E11.101"  
note="probably inactive due to including stop codon(s) in  
CDS

pseudogene, Transposon MAGY gag and pol gene"  
/pseudo  
/complement(10071..11466)  
/gene="OJ1793\_E11.101"  
/complement(10071..11466)  
/gene="OJ1793\_E11.101"  
/note="5' LTR"  
11966..12388  
/gene="OJ1793\_E11.102"  
/note="start and end point are not identified"  
11996..12388  
/gene="OJ1793\_E11.102"  
/note="similar to Oryza sativa chromosome 3,  
OSJNA0081P02.2"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAC19976.1"  
/db\_xref="GI:23495765"  
/translation="MATAVARGSGRDAMGDPPTARSDSDGTARRGLDGTARGAAG  
LEAMERAGDGMATATAROMRARGGTTRRRRGRDQKMTATGSDGGAAGGDDG  
DGRRLATMATRGDCGDKERKEHELG"  
15949..117980  
/gene="OJ1793\_E11.103"  
15949..117980  
/gene="OJ1793\_E11.103"  
/note="probably inactive due to internal exon missing in  
CDS  
pseudogene, Mutator-like transposase"  
/pseudo  
21958..23753  
/gene="OJ1793\_E11.104"  
/join(21958..22287,22350..22451,22581..22790,22934..23014,  
23102..23287,23433..23753)  
/note="OJ1793\_E11.104"  
/note="start and end point are not identified"  
/join(21958..22287,22350..22451,22581..22790,22934..23014,  
23102..23287,23433..23753)  
/gene="OJ1793\_E11.104"  
/codon\_start=1  
/product="putative 5-alpha-taxadienol-10-beta-hydroxylase"  
/protein\_id="BAC19977.1"  
/db\_xref="GI:23495766"  
/translation="WDSMLLALLALFIPILHLVTRRKASYNLPSGSGIFPLIQ  
TISLRARKNDYQYODRIKKYGPVSKSVFSGPTVLTGPANRPAFCNPDLIFT  
OKALNALGYLPEVAVTKYWKMDKESVRHIDLMVQKTLTVAPLAKRLTFNITCY  
PGEBAQPIREALADPVALYKATLSIPVPIPTFKNGLSASNRIRKLRLIYERT  
SVLIIFLRVLANEPIDIGNITBOEIRARNRNPDLMDVSRMKTWKAMETLT  
VPAIFGSPRTAKDIEGQYHFKGMOYFTDQIYTHLDTPNPDGPRKEDPARFNQSS  
IPYCFVPGGSPKCPGNERPAKTIIVAMETLVARQFRMKLCKEGRKDPITMPL  
GLPDLERSPPGVNHS"  
25876..27683  
/gene="OJ1793\_E11.105"  
/join(25876..26358,26486..26924,27017..27204,  
27366..27683)  
/note="start and end point are not identified"  
/join(25876..26358,26486..26924,27017..27204,27366..27683)  
/gene="OJ1793\_E11.105"  
/codon\_start=1  
/product="putative 5-alpha-taxadienol-10-beta-hydroxylase"  
/protein\_id="BAC19978.1"  
/db\_xref="GI:23495767"  
/translation="WPFALLALALPILHLVTRRKASYNLPSGSGIFPVGISIL  
LALNSNDYQYODRIKKYGPVSKSVFSGPTVLTGPANRPAFCNPDLIFTVA  
ANALRSILTYGSELKQYRSALOGYKRTMWKCIKMDSEVRHIDLMVQKTY  
TVAPLAKRLTFNITCYVIFGQAGPIREALADPVALYKATLSIPVPIPTFKNGLS  
ASRKYRLROIARDREAAQOGHSSADDFETVLYRSGSTLSIVEDVDAAIL  
LIAGETSSVITPILRYLANEPIDIGNITBOEIRARNRNPDLMDVSRMKTWK  
KYAMETLTCTPRTISRTAKDIEGQYHFKGMOYFTDQIYTHLDTPNPDGPRKED  
PARFNQSSVPPYCFVPGGSPKCPGNERPAKTIIVAMETLVARQFRMKLCKEGR  
KDPVMPLELPLETRSSPWICLK"

gene 30225..34000  
/gene="OJ1793\_E11.106"  
/join(30225..30404,31022..31115,32139..32530,32639..33077,  
33177..33364,33504..34000)  
/gene="OJ1793\_E11.106"  
/note="supported by full-length cDNA(s) : AK073618"  
/join(32255..32530,32639..33077,33177..33364,33504..33824)  
/gene="OJ1793\_E11.106"  
/note="contains full-length cDNA(s) : AK073618"  
/codon\_start=1  
/product="putative 5-alpha-taxadienol-10-beta-hydroxylase"  
/protein\_id="BAC79651.1"  
/db\_xref="GI:33146518"  
/translation="MSLRGSPVTLAGPGANHFVFSNODLIFTETKAINALVRSILT  
LSGEELKQYRAGALHGLRPEMWTYKMRKMDSEVRHIDLMVQKTYTVAPLAKRL  
DIICSVIRGQVGPPIREALADPVALYKATLSIPVPIPTFKNGLSASNRIRKYLRO  
IARDMGALAQGYSSSADDFETVLYRSGSTLSIVEDVDAAILLIAGETSSVIT  
ITPILRYLANEPIDIGNITBOEIRARNRNPDLMDVSRMKTWKAMETLTIS  
PIGSPRTAKDIEGQYHFKGMOYFTDQIYTHLDTPNPDGPRKEDPARFNQSSILP  
PYCFVPGGSPKCPGNERPAKTIIVAMETLVARQFRMKLCKEGRKDPITPVLGL  
PILETRTPPEVYVHA"  
complement(36390..39069)  
/gene="OJ1793\_E11.107"  
complement(36390..39069)  
/gene="OJ1793\_E11.107"  
/note="probably inactive due to 5' exon missing in CDS  
probably inactive due to including frameshift(s) in CDS  
pseudogene, gag/pol polypeptide"  
/pseudo  
complement(40166..40718)  
/gene="OJ1793\_E11.108"  
complement(40166..40718)  
/gene="OJ1793\_E11.108"  
/note="hypothetical ORF"  
predicted by GENSCAN  
this category is not included in IRGSP standard"  
complement(41589..41794,41957..42017)  
/gene="OJ1793\_E11.109"  
complement(41589..41794,41957..42017)  
/gene="OJ1793\_E11.109"  
/note="hypothetical ORF"  
predicted by GENSCAN  
this category is not included in IRGSP standard"  
43397..45180  
/gene="OJ1793\_E11.110"  
/join(43397..43891,43997..44435,44534..44721,  
44875..45180)  
Query Match 6.5%; Score 36; DB 8; Length 101342;  
Best Local Similarity 62.0%; Pred. No. 13;  
Matches 57; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
gene 43397..45180  
mRNA 77729 CCCTTAAGGGTCGCTTGCAAAATAGATTTCGTAGCAGCAACGAGCTACCCCA 77670  
QY 52 CCCTTAAGTCATGCTGTGCAATAATGGTATCGTACTCAACATTAGTCAACTCA 111  
Db 77729 CCCTTAAGGGTCGCTTGCAAAATAGATTTCGTAGCAGCAACGAGCTACCCCA 77670  
QY 112 CTCGCTATCGATGATGAGCCATTTCGCCG 143  
Db 77669 GTTATATTTTGAAGCGGCTATTATTCGCG 77638  
RESULT 13  
AC112046/c 271370 bp DNA linear HTG 10-OCT-2002  
LOCUS AC112046  
DEFINITION Rattus norvegicus clone CH230-71F9, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
3 unordered pieces.  
ACCESSION AC112046  
VERSION AC112046.3 GI:23270281  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;



```

ORGANISM      Caenorhabditis elegans
REFERENCE     Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida;
AUTHORS      Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
TITLE        1
JOURNAL      none.
MEDLINE      Genome sequence of the nematode C. elegans: a platform for
REFERENCE     investigating biology. The C. elegans Sequencing Consortium
AUTHORS      Science 282 (5396), 2012-2018 (1998)
JOURNAL      99069613
TITLE        The C. elegans Sequencing Consortium.
REFERENCE     2 (bases 1 to 27243)
AUTHORS      Ainscough, R.
JOURNAL      Direct Submission
TITLE        Submitted (09-AUG-1996) Nematode Sequencing Project, Sanger
REFERENCE     Institute, Hinxton, Cambridge CB10 1SA, England and Department of
AUTHORS      Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
JOURNAL      jee@sanger.ac.uk or rwenematode@wustl.edu
COMMENT       Coding sequences below are predicted from computer analysis, using
              predictions from GeneFinder (P. Green, U. Washington), and other
              available information.
              Current sequence finishing criteria for the C. elegans genome
              sequencing consortium are that all bases are either sequenced
              unambiguously on both strands, or on a single strand with both a
              dye primer and dye terminator reaction, from distinct subclones.
              Exceptions are indicated by an explicit note.
              For a graphical representation of this sequence and its analysis
              see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
              name=W06A7
              IMPORTANT: This sequence is NOT necessarily the entire insert of
              the specified clone. It may be shorter because we only sequence
              overlapping sections once, or longer because we arrange for a small
              overlap between neighbouring submissions.
              IMPORTANT: This sequence is not the entire insert of clone W06A7.
              It may be shorter because we only sequence overlapping sections
              once, or longer because we arrange for a small overlap between
              neighbouring submissions.
              The true left end of clone W06A7 is at 1 in this sequence. The true
              right end of clone W06A7 is at 9975 in
              sequence 278067.
              The true left end of clone ZC412 is at 27138 in this sequence. The
              true right end of clone C48G7 is at 10917 in this sequence. The
              start of this sequence (1..103) overlaps with the end of sequence
              278061.
              The end of this sequence (27138..27243) overlaps with the start of
              sequence 278067.
              Location/Qualifiers
                1..27243
                /organism="Caenorhabditis elegans"
                /mol_type="Genomic DNA"
                /strain="Bristol N2"
                /db_xref="taxon:6239"
                /chromosome="IV"
                /clone="W06A7"
                /join(4610..4719,4772..4870,5394..5604,5655..5725,
                /gene="W06A7.2"
                /join(4610..4719,4772..4870,5394..5604,5655..5725,
                /gene="W06A7.2"
                /join(4610..4719,4772..4870,5394..5604,5655..5725,
                /gene="W06A7.2"
                /standard_name="W06A7.2"
                /codon_start=1
                /product="Hypothetical protein W06A7.2"
                /protein_id="CA801521.1"
                /db_xref="GI:3880493"
                /translation="MKRTVDGVVPPVSRAPVPCPLPAMGIMWPAKQVLTIG
                ALTFPFGVITLMLTCTAIECGAVASALIPALITITIGTLLFCGCAAHLDHNS
                EIPFKITTTTITVMDPESHEDIRLFTVHRBQVFNWQSGWQPDENLSWQASNP
                YPIQHTXNCLERQPY"
                complement(9699..9803)
                /gene="W06A7.5"
                /complement(9699..9803)
                /gene="W06A7.5"
                /standard_name="W06A7.5"

```



Query Match	6.5%; Score 35.8; DB 3; Length 27243;
Best Local Similarity	50.3%; Pred. No. 12;
Matches	88; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 277 TCACCTTGCAAGAGAGCTTCATGTCATTAATTTAGCAATTTAGGTTAGGGTGCA	336
Db 369 TTTCATTTTAAACATTAATTTTCAATTAATCTGATTAATAGTTATGSTATAGTC	428
QY 337 CTAGTAATTCAGCAGAAATTTGGGTTAAAGACACACACAGATTCGCGTATCCACA	396
Db 429 AATGGAATATATATCTTAATCTATAGTTATGAAATAATATACACAAAAATCTGACACAC	488
QY 397 AATTCAGGACGATTTGCGAGACATTAACCTTAGGGCGGCGACCCCTTTACTGC	451
Db 489 AATTATTAAGGCTGACACCGAGACATTCAGAGGCACTTCATTAATTTTGGC	543

RESULT 15

AC069562

LOCUS

DEFINITION

Mus musculus clone rp23-317e13 map 13 strain C57BL/6J, complete

AC069562 215278 bp DNA linear ROD 18-SEP-2003

AC069562

VERSION AC069562.51 GI:34849934

```

KEYWORDS      HTG.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
TITLE         1 (bases 1 to 215278)
JOURNAL       Song,L., Jiang,X., Swank,R. and Roe,B.A.
REFERENCE      Mus musculus Chromosome 13 BAC Clone rp23-317e13
AUTHORS       Unpublished
TITLE         2 (bases 1 to 215278)
JOURNAL       Song,L., Jiang,X., Swank,R. and Roe,B.A.
REFERENCE      Direct Submission
AUTHORS       Submitted (03-JUN-2000) Department Of Chemistry And Biochemistry,
TITLE         The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
JOURNAL       OK 73019, USA
REFERENCE      3 (bases 1 to 215278)
AUTHORS       Song,L., Jiang,X., Swank,R. and Roe,B.A.
TITLE         Direct Submission
JOURNAL       Submitted (11-JUN-2003) Department Of Chemistry And Biochemistry,
               The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
               OK 73019, USA
REFERENCE      4 (bases 1 to 215278)
AUTHORS       Song,L., Jiang,X., Swank,R. and Roe,B.A.
TITLE         Direct Submission
JOURNAL       Submitted (20-JUN-2003) Department Of Chemistry And Biochemistry,
               The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
               OK 73019, USA
REFERENCE      5 (bases 1 to 215278)
AUTHORS       Song,L., Jiang,X., Swank,R. and Roe,B.A.
TITLE         Direct Submission
JOURNAL       Submitted (24-JUN-2003) Department Of Chemistry And Biochemistry,
               The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
               OK 73019, USA
REFERENCE      6 (bases 1 to 215278)
AUTHORS       Song,L., Jiang,X., Swank,R. and Roe,B.A.
TITLE         Direct Submission
JOURNAL       Submitted (18-SEP-2003) Department Of Chemistry And Biochemistry,
               The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
               OK 73019, USA
COMMENT       On Sep 18, 2003 this sequence version replaced gi:32171311.
               ----- Genome Center
               Center: Department Of Chemistry And Biochemistry
               The University Of Oklahoma
               Center code:UOKNOR
               -----
FEATURES             Location/Qualifiers
     source           1. 215278
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="13"
                     /clone="rp23-317e13"
                     /clone_1ib="RPCT - 23 Female (C57BL/6J) Mouse BAC Library"
ORIGIN
Query Match      6.5%; Score 35.8; DB 10; Length 215278;
Best Local Similarity 52.3%; Pred. No. 17;
Matches          79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
249 AACCTCTCCCATGTCGTGCACATCGGCTCACCTTGGCAAGAGCTTCATTGTCAATA 308
DB      78575 AAACAGTACACGTAAGTAACATCAGCAAAACGCTGTGCAGAACTATCTTGTGAACTG 7863
QY      309 AATTGACCAATTAGCTTTAGGGGTGCGACTAGTAATTACGACGAATTTGGGTTAAAG 368
DB      78635 CATTAAAGAAATGTTTCCCTCGAAGGAAAACTGAATTCATGAGACTTGGGTCAAGAT 7869
QY      369 ACACAGACAGCAATCCGCTATCCACAAAA 399
DB      78695 AGATGTAAATGAATCCAGTAACCACTTATA 78725

```



Mon May 3 09:19:49 2004

us-10-603-260-4.rge

Page 19

Search completed: April 30, 2004, 15:29:14  
Job time : 2249.64 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 30, 2004, 06:19:01 ; Search time 227.167 Seconds  
(without alignments)  
10378.910 Million cell updates/sec

Title: US-10-603-260-4

Perfect score: 555

Sequence: 1 atgcagatattcttcttca.....cttcatacagaatgcgtt 555

Scoring table: IDENTITY NUC

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*\n2: Geneseq1990s:\*\n3: Geneseq2000s:\*\n4: Geneseq2001as:\*\n5: Geneseq2001bs:\*\n6: Geneseq2002as:\*\n7: Geneseq2003as:\*\n8: Geneseq2003bs:\*\n9: Geneseq2003cs:\*\n10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	13.7	605	2	AAT79331
2	37.2	6.7	2000	7	ADA71938
3	36.8	6.6	26997	4	AA646747
4	35.6	6.4	115218	7	ACA64845
5	34.4	6.2	2277	2	AAV13834
6	34.4	6.2	2277	2	AAV05370
7	33.2	6.0	2000	7	ADA71938
8	32.8	5.9	948	7	ACA27498
9	32.4	5.8	91080	6	ABX08336_16
10	32.4	5.8	110000	6	ABX08336_15
11	32	5.8	549	7	ACA30202
12	32	5.8	936	4	AAH31609
13	31.8	5.7	528	7	AAH73424
14	31.8	5.7	969	6	ABK65284
15	31.8	5.7	969	6	ABK65284
16	31.8	5.7	969	6	ABK65284
17	31.8	5.7	969	6	ABK65284
18	31.8	5.7	4715	2	AAV74514
19	31.6	5.7	409	4	AAH83491
20	31.6	5.7	2150	6	ABK85746
21	31.6	5.7	2217	9	ADC92937
22	31.6	5.7	17286	4	AAH36869
23	31.6	5.7	17286	7	ABX59857

24	31.4	5.7	2058	5	AA693936	AA693936 DNA encod
25	31.2	5.6	1300	6	ABK65274	ABK65274 Arabidops
26	31.2	5.6	1300	7	ADD30808	ADD30808 Plant yle
27	31.2	5.6	1308	7	ACA26237	ACA26237 Prokaryot
28	31.2	5.6	11485	9	AD879904	AD879904 Mouse put
29	31.2	5.6	15275	9	AA635975	AA635975 Human car
30	31.2	5.6	15275	9	AD846669	AD846669 Human car
31	31	5.6	918	2	AAH28551	AAH28551 Bacterial
32	31	5.6	918	2	AAH28551	AAH28551 Bacterial
33	31	5.6	4590	5	AAH24065	AAH24065 Yeast AD
34	31	5.6	5162	2	AAV69919	AAV69919 Expressio
35	31	5.6	5162	2	AAV69922	AAV69922 Multigenic
36	31	5.6	5262	2	AAV69927	AAV69927 Promoter
37	31	5.6	6950	2	AAV69920	AAV69920 Expressio
38	31	5.6	105194	6	ABK24132	ABK24132 Bacterial
39	30.8	5.5	33248	4	AAH28537	AAH28537 Genomic f
40	30.6	5.5	476	4	AAH15636	AAH15636 Probe #56
41	30.6	5.5	476	4	ABK57784	ABK57784 Human fce
42	30.6	5.5	476	4	AAH37368	AAH37368 Probe #60
43	30.6	5.5	476	4	ABK27146	ABK27146 Probe #56
44	30.6	5.5	476	4	AAK31484	AAK31484 Human bon
45	30.6	5.5	476	4	AAK05851	AAK05851 Human bra

## ALIGNMENTS

RESULT 1	AAH79331	standard, DNA, 605 BP.
AC	AAH79331,	
XX	27-AUG-2003 (revised)	
DT	16-FEB-1998 (first entry)	
XX	DNA encoding L11.1 esterase es2.	
XX	Esterase, thermostable enzyme; ester; chiral compound; cheese; pulp;	
KW	paper; lignin removal; sugar; lignocellulose; disease resistance;	
KW	feedstuff; ss.	
XX	Unidentified.	
OS	WO9730160-A1.	
PN	21-AUG-1997.	
PD	11-FEB-1997; 97WO-US002039.	
XX	16-FEB-1996; 96US-00602359.	
PR	(RECO-) RECOMBINANT BIOCATALYSIS INC.	
PA	Robertson DE, Murphy D, Reid J, Maffia AM, Link S, Swanson RV;	
XX	Warren PV, Koehnocka A, Callen W;	
PI	WPI, 1997-425035/39.	
XX	P-PSDB; AAW23078.	
DR	Nucleic acid encoding heat stable esterase from thermophilic bacteria -	
XX	which is active in organic solvents, useful in cheese or paper	
PT	manufacture, and to study plant resistance to disease.	
PT	Disclosure, Page 63-64, 113pp; English.	
PS	This DNA sequence codes for L11.1 esterase es2 (AAW23078). Newly	
XX	identified polynucleotides (AAH79331-40) encoding esterases (AAW23069-	
CC	88), some of which are claimed, can be used for recombinant production of	
CC	the enzymes in host cells, and as probes to identify related sequences.	
CC	The esterases are stable at high temperature and in organic solvents,	
CC	making them superior for use in production of pure chiral compounds used	
CC	in pharmaceutical, agricultural and other chemical industries. A method	



XX AAC46747;  
AC  
XX  
DT 18-DEC-2001 (first entry)  
DE Tumour suppressor gene derived chemically modified sequence #471.  
XX  
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;  
KM tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
XX cytosine methylation; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200168912-A2.  
XX  
XX 20-SEP-2001.  
XX  
XX 15-MAR-2001; 2001WO-EP002955.  
XX  
XX 15-MAR-2000; 2000DE-01013847.  
XX 06-APR-2000; 2000DE-01019058.  
XX 07-APR-2000; 2000DE-01019173.  
XX 30-JUN-2000; 2000DE-01032529.  
XX 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX MPI; 2001-602752/68.  
XX  
XX  
XX Fragments of chemically modified genes associated with tumor suppressor  
XX genes and oncogenes, useful in designing primers and probes for analyzing  
XX diseases associated with cytosine methylation state e.g. cancer.  
XX  
XX  
XX Claim 1; SEQ ID NO 471; 27pp; English.  
XX  
XX  
XX The invention relates to a nucleic acid comprising a sequence of 18  
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
XX bisulphite, of genes associated with tumour suppression and oncogenes  
XX having a sequence taken from 536 (actually 533 since numbers 408, 458 and  
XX 500 are missing from the sequence listing) sequences (Ss) and sequences  
XX complementary to (Ss). The nucleic acid may be a peptide nucleic acid-  
XX oligomer (PNA) of at least 9 nucleotides and may form part of a set of  
XX probes for detecting the cytosine methylation state and/or single  
XX nucleotide polymorphisms and also to be used in an array for analysing  
XX diseases associated with CpG dinucleotides e.g. cancers and tumours. The  
XX probes can also be used in a method for ascertaining genetic and/or  
XX epigenetic parameters for the diagnosis and/or therapy of existing  
XX diseases or the predisposition to specific diseases, by analysing  
XX cytosine methylations. The parameters may be compared to another set of  
XX genetic and/or epigenetic parameters, the differences serving as basis  
XX for diagnosis and/or prognosis events which are disadvantageous to  
XX patients. The present sequence is one of the 533 genomic sequences  
XX derived from tumour suppressor genes and oncogenes. Note: the sequence  
XX data for this patent did not form part of the printed specification, but  
XX was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 26997 BP; 7233 A; 354 C; 6190 G; 13220 T; 0 U; 0 Other;  
SQ

Query Match 6.6%; Score 36.8; DB 4; Length 26997;  
Best Local Similarity 47.1%; Pred. No. 0.57; Indels 0; Gaps 0;  
Matches 113; Conservative 0; Mismatches 127;

QY 277 TCACCTTGCAAGAGCTTCATTCATATAAATTGACATTTAGGTTAGGGGTGCA 336  
DB 13416 TCTCATGAAAGTAACTAACATCCATATACAAATTAACCTAAATACGTTAATTC 13357  
QY 337 CTAGGTAATTCAGCAAAATTGGGTTAAAGAACGACGACGAATCCCGCTATCCACA 396  
DB 13356 CTACCAACATTTATCAACTTCGACACTAATATACACTGAAACCAATCATCTTTAAACAA 13297

QY 397 AAATCAGGCGATTTGCAGAACATACCTTTAGGGCTGCGACCTTTACTGCGCAT 456  
DB 13296 ATAAATATATAAACTACAAAACGCTACCTTATATCTACAAATATCCCGACTTAC 13237  
QY 457 CCAGTGAAGTCCGATGATGATCCGTCACAGTAGAAGAAACAAATATGCGCATGACAGAT 516  
DB 13236 CCACCAACGACCAATACACTCTCCCACTTATACCAACAAACGAAACATTACAAAT 13177

RESULT 4  
ACA64845  
ID ACA64845 standard; DNA; 115218 BP.  
XX  
XX ACA64845;  
XX  
XX 27-JUN-2003 (first entry)  
XX  
XX Human HNRP GP43 DNA corresponding to AL034397.  
XX  
XX Human; chronic inflammatory joint disease; infection; tumour;  
XX antiinflammatory; cytostatic; antiarthritic; antirheumatic;  
XX immunosuppressive; gene therapy; etiological pathogenicity; ds.  
XX  
XX Homo sapiens.  
XX  
XX DE10127572-AL.  
XX  
XX 05-DEC-2002.  
XX  
XX 30-MAY-2001; 2001DE-01021572.  
XX  
XX 30-MAY-2001; 2001DE-01027572.  
XX  
XX (PAT-) PATHOARRAY GMBH.  
XX  
XX  
XX Haep1 T, Ungethlem U, Blass S;  
XX  
XX MPI; 2003-240797/24.  
XX  
XX  
XX Reagents for diagnosis, study and therapy of chronic inflammatory joint  
XX PT and other diseases, comprises any of many specified genes or derived  
XX PT proteins.  
XX  
XX Claim 1; Page; 12pp; German.  
XX  
XX This invention describes a novel reagent for diagnosis, molecular  
XX definition and therapy of chronic inflammatory joint diseases, and other  
XX CC inflammatory disorders, infective or tumour diseases in humans. The  
XX CC products of the invention have antiinflammatory, cytostatic,  
XX CC antiarthritic, antirheumatic and immunosuppressive activity and can be  
XX CC used for gene therapy. The reagent of the invention and any proteins and  
XX CC antibodies derived from it, are used (i) for analysing tissue and blood  
XX CC samples for medical diagnosis; (ii) for diagnosis and characterisation of  
XX CC chronic joint diseases, on the basis of molecular characterisation, and  
XX CC determining the etiological pathogenicity principle of as yet  
XX CC uncharacterised inflammatory diseases, also monitoring progression and/or  
XX CC treatment of disease, and optimisation of therapy and (iii) for  
XX CC developing treatments for inflammatory diseases, particularly of joints,  
XX CC infections and tumours. ACA64801-ACA64965 represent human polymucleotides  
XX CC used in the method of the invention  
XX  
XX Sequence 115218 BP; 35618 A; 23477 C; 22286 G; 33837 T; 0 U; 0 Other;  
SQ

Query Match 6.4%; Score 35.6; DB 7; Length 115218;  
Best Local Similarity 52.7%; Pred. No. 3.1; Indels 0; Gaps 0;  
Matches 77; Conservative 0; Mismatches 69;

QY 1 ATGCACATATCTTCTGCATGACCTATATGACATGGCTGTATGATCGCTTACT 60  
DB 53509 ATACACTTCTCTCTTTAGGGAATCAATTTAAGAGGTTGGATTAACATGTGCTAGT 53568  
QY 61 CATGCTGCATATAATTTGGGTTATGCTACTCAACCATTTAGCTACACTCGCTATC 120

DB 53569 TCTTATCAAGATAAATACTATTTCATGGGAAAGCCCTTGGTCTACTGAAATCCCTTC 53628  
 QY 121 GATGATGAGGCCATTTTTCGCCGCT 146  
 DB 53629 CATGGTAAAGAGAGTGGTGAAGCT 53654

RESULT 5  
 AAV13834/c  
 ID AAV13834 standard; cDNA; 2277 BP.  
 XX  
 AC AAV13834;  
 DT 21-JUL-1998 (first entry)  
 DE Homo sapiens ambiguity-maximised telomerase protein p105 gene.  
 XX  
 XX telomerase; p105; treatment; prevention; cancer; restenosis;  
 XX inflammation; myocardial infarction; glomerulonephritis; transplant;  
 XX rejection; infection; HIV; human immunodeficiency virus;  
 XX bone marrow transplants; proliferation-restricted cells;  
 XX ambiguity-maximised; ds.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT 1. 2277  
 FT CDS /\*tag= a  
 FT /product= "telomerase protein p105"  
 XX  
 XX WO9801543-A1.  
 XX 15-JAN-1998.  
 XX  
 XX 08-JUL-1997; 97WO-US012297.  
 XX  
 XX 08-JUL-1996; 96US-00676967.  
 XX  
 XX (TULUA-) TULARIK INC.  
 XX  
 XX Cao Z;  
 XX  
 XX WPI; 1998-101044/09.  
 XX P-PSDB; AAW41927.  
 DR  
 XX  
 PT New nucleic acid encoding human telomerase protein p105 or its fragments  
 PT - used for therapeutic modulation of telomerase activity and for  
 PT screening for potential modulators of telomerase-target binding.  
 XX  
 XX Disclosure; Page 20-21; 32pp; English.  
 XX  
 XX The sequence is that of an ambiguity-maximised human telomerase protein  
 XX coding sequence. The sequence, or specific fragments of it, can be used  
 XX to modulate expression of a telomerase transcript (by hybridising to it  
 XX intracellularly), e.g. for treatment or prevention of cancer, restenosis,  
 XX inflammation, myocardial infarction, glomerulonephritis, transplant  
 XX rejection and infections (e.g. with human immunodeficiency virus). It can  
 XX be used to express recombinant telomerase protein which can be used to  
 XX screen for agents, e.g. antibodies, that modulate binding of human  
 XX telomerase to its binding target. Those that inhibit telomerase activity  
 XX can be used to treat the conditions listed above, while those that are  
 XX agonists can be used to extend the life of proliferation-restricted  
 XX cells, especially normal somatic cells, e.g. in cases of hypersensitivity  
 XX or atrophy, also to improve production of recombinant proteins by  
 XX maximising cell density and survival and expansion of precursor cells  
 XX being used for bone marrow transplants. They may also be used for  
 XX diagnosis. Other uses of telomerase proteins are isolation, enrichment  
 XX and concentration of telomerase RNA or proteins; as immunogens; in  
 XX therapy, as reagent where nascent oligonucleotides of known structure are  
 XX needed (e.g. for tagging native nucleic acid molecules) and for  
 XX regulating cell growth/density tolerance. The agents and the telomerase  
 XX proteins should be very specific, e.g. they are selective for cancer

CC cells without harming somatic cells  
 XX  
 SQ Sequence 2277 BP, 513 A, 212 C, 395 G, 217 T, 0 U, 940 Other;  
 Query Match 6.2%; Score 34.4; DB 2; Length 2277;  
 Best Local Similarity 34.5%; Pred. No. 1.1;  
 Matches 51; Conservative 25; Mismatches 72; Indels 0; Gaps 0;

QY 11 TCTGTTCAATGACCTCTATATGATGAGCTTGGTAAATGATCCGCTATGATCTCTCC 70  
 DB 662 TTTTYYTNACNSWYCYGRTGYTNNSWYTCGNSWYTTTCYCNCCDATTNGNSWN 603  
 QY 71 ATAAATGGGTATCGACTCAACACATTAAGTACACTCACTCGCTATCGATGAGAG 130  
 DB 602 ACNSWYTGAGTCTCTTATATTTCTTTNGCAGACGCCATRCANACNGMANGTNGX 543  
 QY 131 CCAATTTTCGCCGCTTGACCGATGCT 158  
 DB 542 CCYTDTATYTCYTTCAATRTTCATNCCYT 515

RESULT 6  
 AAV05370/c  
 ID AAV05370 standard; cDNA; 2277 BP.  
 XX  
 AC AAV05370;  
 DT 06-JUL-1998 (first entry)  
 DE Homo sapiens telomerase p105 subunit synthetic gene.  
 XX  
 XX Telomerase; p105; human; cell replication; cancer; restenosis;  
 XX multiple sclerosis; inflammation; rheumatoid arthritis;  
 XX myocardial infarction; glomerulonephritis; transplant rejection;  
 XX infection; therapy; ds.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX WO9801542-A1.  
 XX 15-JAN-1998.  
 XX  
 XX 08-JUL-1997; 97WO-US012296.  
 XX  
 XX 08-JUL-1996; 96US-00676974.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Collins K;  
 XX  
 XX WPI; 1998-101043/09.  
 DR  
 XX  
 PT New nucleic acid encoding human telomerase proteins or their fragments -  
 PT useful for therapeutic modulation of telomerase activity and for  
 PT screening for potential modulators of telomerase-target binding.  
 XX  
 XX Disclosure; Page 19-20; 32pp; English.  
 XX  
 XX This polynucleotide comprises a synthetic, ambiguity-maximised DNA coding  
 XX for the p105 subunit (see AAW46593) of human telomerase. It is based on  
 XX an isolated cDNA clone (see AAV05369) for p105 and encompasses all  
 XX possible nucleic acids encoding the full-length protein. The invention  
 XX provides methods relating to human telomerase and related nucleic acids,  
 XX including the subunit proteins p140, p105, p48 and p43. The proteins may  
 XX be produced recombinantly from transformed host cells or purified from  
 XX human cells. Also included are human telomerase RNA (see AAV05373) and  
 XX functional derivatives (see AAV05374 and AAV16092-93), as well as p105  
 XX synthetic DNA sequences (AAV05370-72). The invention also provides  
 XX isolated telomerase hybridisation probes and primers capable of  
 XX specifically hybridising with the telomerase gene, telomerase-specific  
 XX binding agents such as specific antibodies, and methods of making and  
 XX using the subject compositions in diagnosis (e.g. genetic hybridisation



PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 14; SEQ ID NO 15368; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation; or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences

XX Sequence 948 BP; 370 A; 114 C; 204 G; 260 T; 0 U; 0 Other;

Query Match 5.8%; Score 32.8; DB 7; Length 948;

Best Local Similarity 46.8%; Pred. No. 2.6; Mismatches 117; Indels 0; Gaps 0;

Matches 103; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 336 ACTAGTATTGACGAGAAATTTGGTTAAAGAACGACGAGCAATCCCGCTATCCACA 395  
 DB 561 AATACTGATTACGACAACTACCACTAGAGATGAGGAAATGTACTATTAAGACCA 620  
 QY 396 AAATCGAGCATTTGCAAGAACGATACCTTAGGGCTGCGAGCCTTTACTGGCGCA 455  
 DB 621 AGAAGCATTTAGTTGGCAGCAAGAACTTATAGCAATTTTAAATTTTATGACTCT 680  
 QY 456 TCCACTGAGCTCCGATGTAACGCTCAAGTAAAGAAACCAAAATAGCTGATGACAGA 515  
 DB 681 TAGCATCATGACGATATGTGAGATTAATGTAGAAAAGAAAGATAGAGAGAAA 740  
 QY 516 TCATATGCGATATCCACCACTTCATACGAGATGCTGTT 555  
 DB 741 GGTCTTGAGATGACTATCGAAGAACTTATCTTCACTT 780

# RESULT 9

ABX08336\_16/c

Continuation (17 of 17) of ABX08336 from base 160001 (Human phosphodiesterase 4D (PDE4D

WP Sequence split info 17 fragments LOCUS ABX08336 Accession ABX08336

Fragment Name	Begin	End
WP ABX08336_00	1	110000
WP ABX08336_01	100001	210000
WP ABX08336_02	200001	310000
WP ABX08336_03	300001	410000
WP ABX08336_04	400001	510000
WP ABX08336_05	500001	610000
WP ABX08336_06	600001	710000
WP ABX08336_07	700001	810000

WP ABX08336\_08 800001 910000  
 WP ABX08336\_09 900001 1010000  
 WP ABX08336\_10 100001 1110000  
 WP ABX08336\_11 110001 1210000  
 WP ABX08336\_12 120001 1310000  
 WP ABX08336\_13 130001 1410000  
 WP ABX08336\_14 140001 1510000  
 WP ABX08336\_15 150001 1610000  
 WP ABX08336\_16 160001 1691080

Query Match 5.8%; Score 32.4; DB 6; Length 91080;

Best Local Similarity 50.0%; Pred. No. 35; Mismatches 81; Indels 0; Gaps 0;

Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 274 GGCTCACCCTTGGACAGAGGCTTCATGCAATTAATGCAATTAGGGGTG 333  
 DB 2603 GGCTCATCTTACATATTTCTCTGAGTCCCTAGAAATGCTATTTTTCAAAAG 2544  
 QY 334 GCACTAGTAAATTCAGAGAAATTTGGTTAAAGAACGACGACGAAATCCGCTATCCA 393  
 DB 2543 TCTTGTTCTTTTACAGAGAAATGTATAGAAACCAAAATGAGGTGCTAAGTGCT 2484  
 QY 394 CAAAATCAGGCAATTTGACAGAAATCCTTTAGGGGTG 435  
 DB 2483 CATTGCTACTGGGTATGTAGCTTCTAGACCTCTCGGCTG 2442

# RESULT 10

ABX08336\_15/c

Continuation (16 of 17) of ABX08336 from base 150001 (Human phosphodiesterase 4D (PDE4

WP Sequence split info 17 fragments LOCUS ABX08336 Accession ABX08336

Fragment Name	Begin	End
WP ABX08336_00	1	110000
WP ABX08336_01	100001	210000
WP ABX08336_02	200001	310000
WP ABX08336_03	300001	410000
WP ABX08336_04	400001	510000
WP ABX08336_05	500001	610000
WP ABX08336_06	600001	710000
WP ABX08336_07	700001	810000
WP ABX08336_08	800001	910000
WP ABX08336_09	900001	1010000
WP ABX08336_10	1000001	1110000
WP ABX08336_11	1100001	1210000
WP ABX08336_12	1200001	1310000
WP ABX08336_13	1300001	1410000
WP ABX08336_14	1400001	1510000
WP ABX08336_15	1500001	1610000
WP ABX08336_16	1600001	1691080

Query Match 5.8%; Score 32.4; DB 6; Length 110000;

Best Local Similarity 50.0%; Pred. No. 39; Mismatches 81; Indels 0; Gaps 0;

Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 274 GGCTCACCCTTGGACAGAGCTTCATGCAATTAATGCAATTAGGGGTG 333  
 DB 102603 GGCTCATCTTACATATTTCTCTGAGTCCCTAGAAATGCTATTTTTCAAAAG 102544  
 QY 334 GCACTAGTAAATTCAGAGAAATTTGGTTAAAGAACGACGACGAAATCCGCTATCCA 393  
 DB 102543 TCTTGTTCTTTTACAGAGAAATGTATAGAAACCAAAATGAGGTGCTAAGTGCT 102484  
 QY 394 CAAAATCAGGCAATTTGACAGAAATCCTTTAGGGGTG 435  
 DB 102483 CATTGCTACTGGGTATGTAGCTTCTAGACCTCTCGGCTG 102442

# RESULT 11

ACA30202

ID ACA30202 standard; DNA; 549 BP.



DT 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #1859.  
DE  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KM drug design; gene.  
XX  
XX Campylobacter jejuni.  
OS  
XX WO200277183-A2.  
XX  
XX  
XX 03-OCT-2002.  
XX  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948893.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselebeck R, Ohlsen KL, Zyskind JM,  
P1 Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth FA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU26332.  
XX  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids, required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX  
XX Claim 14; SEQ ID NO 18072; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences  
XX  
XX Sequence 549 bp; 198 A; 64 C; 116 G; 171 T; 0 U; 0 Other;

Qy	332	TGGCATTAGTATATCAGAGAA	TTGGTTAAAAAGACGACGACGAAATCCGGATC	391
Db	428	TGATATTGGAAATCCTGCAAAATTTG	CCGGAATTAAAGATGAGAAATCAGCTTC	487
Qy	392	CACAAATATCAG	403	
Db	488	TAAATCAATCCG	499	

RESULT 12  
AAH31609  
ID AAH31609 standard; DNA; 936 BP.

DT 30-JUL-2001 (first entry)  
XX  
DE Human olfactory receptor polynucleotide, SEQ ID NO: 182.

KM Human; olfactory receptor; OR; primary scent determination;  
KM secondary scent determination; polypeptide library; odour receptor  
KM scent profile; scent fingerprint; scent representation; ds.

**Os Homo sapiens.**

PN WO200127158-A2

PD 19-APR-2001

PF 06-OCT-2000; 2000WO-US027582.

PR 08-OCT-1999; 99US-0158615P.

XXXXXX

PA (YEDA ) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Glusman G, Fuchs T, Yanai I,

DR WPI; 2001-290713/30.

**PT** New polynucleotides which encode polypeptides involved in olfactory

XX  
PS C:\a\j\m 8: Page 248: 1857nn, English

The present sequence is one of a number of isolated polynucleotides which

can be used in screening for olfactory agonists and antagonists. The

identification of the odour receptors used to detect these primary

identification of odour receptors that are involved in

representation (also called a scent fingerprint or scent profile), which

CC receptors are useful for determining the interaction pattern of a

CC differences in the olfactory faculties of different individuals

**SQ** Sequence 936 BP; 214 A; 232 C; 174 G; 316 T; 0 U; 0 Other;

Query match	5.8%	Score 32	DB 4	Length 936
-------------	------	----------	------	------------

Matches	50;	Conservative	0;	Mismatches	30;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

259 CATGTCGTCGCATCGGCTCACCTTTGCAAGAGCTTCCATTGTCAATAAATTGAGCAA 318

Db 358 CGTTTCATTGCCATCCGCAATCCTTGAGATATGCTTCCATTTTCAACCAATACTAGAGTC 417

319 TTAGGTTAGGGGTGGCACT 338

Db 418 ATAGCGTTAGGAGTGGGAGT 437

RESULT 13  
ACF73424/C  
ID ACF73424 standard; DNA; 528 BP.  
XX  
XX ACF73424;  
AC  
XX 20-NOV-2003 (first entry)  
XX  
XX Staphylococcus aureus DNA #1104.  
DE  
XX Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
XX enzymatic assay; antibiotic target; gene; ds.  
XX  
XX Staphylococcus aureus.  
OS  
XX WO200294868-A2.  
PN  
XX 28-NOV-2002.  
XX  
XX 27-MAR-2002; 2002WO-1B002637.  
PF  
XX 27-MAR-2001; 2001GB-00007661.  
PR  
XX (CHIR-) CHIRON SPA.  
PA  
XX Masignani V, Mora M, Scarselli M,  
PI  
XX WPI; 2003-120786/11.  
DR  
XX P-PSDB; ABW71864.  
XX  
XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
PT preventing Staphylococcal infection, specifically an infection caused by  
PT S. aureus, e.g. sepsis.  
XX  
XX Claim 6; SEQ ID NO 2207; 49pp; English.  
PS  
XX The invention relates to novel genes and encoded proteins from  
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a  
CC nucleic acid encoding the protein, or an antibody to the protein, is  
CC useful as a pharmaceutical, particularly as a vaccine for treating or  
CC preventing infection due to Staphylococcus bacteria, specifically an  
CC infection caused by S. aureus. The composition is particularly useful for  
CC treating or preventing sepsis in a patient. The composition can also be  
CC used for diagnostics. The protein is also used in an assay for enzymatic  
CC studies and as a target for antibiotics. This sequence represents one of  
CC the novel S. aureus genes of the invention  
XX  
XX Sequence 528 BP; 212 A; 57 C; 116 G; 143 T; 0 U; 0 Other;  
SQ  
Query Match 5.7%; Score 31.8; DB 7; Length 528;  
Best Local Similarity 59.3%; Pred. No. 4.3;  
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
QY 2 TGGAGTTTCTGTTGTTGATGACCTATATGATGCTGCTGATGATCGCTTATGTC 61  
DB 480 TTTATTTTATCTCTTCTTCAATGATATCCGATGATGATTTCTTCTACCTGCTC 421  
QY 62 ATCGTCTGATTAATGGCTTATCGTACTCA 92  
DB 420 ATCTTCTGATATTTAGTTCTTCTTGATCA 390  
RESULT 14  
ABK65284  
ID ABK65284 standard; CDNA; 969 BP.  
XX  
XX ABK65284;  
AC  
XX 02-UTL-2002 (first entry)  
XX  
XX Arabidopsis cDNA encoding a transcription factor #136.  
DE

XX  
XX Plant; ss; gene; transcription factor; transgenic; agriculture;  
XX metabolic chemical; environmental stress; drought;  
XX microbial disease resistance; herbicide resistance; seed yield;  
XX fruit yield; growth rate; leaf senescence; flower senescence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX NO200215675-A1.  
PN  
XX 28-FEB-2002.  
XX  
XX 22-AUG-2001; 2001WO-US026189.  
PF  
XX 22-AUG-2000; 2000US-0227439P.  
PR 16-NOV-2000; 2000US-0071994.  
PR 18-APR-2001; 2001US-00837944.  
XX  
XX (MENDEL-) MENDEL BIOTECHNOLOGY INC.  
PA (PILG/) PILGRIM M.  
PA (CREE/) CREELMAN R.  
PA (DUBE/) DUBELL A J.  
PA (HEAR/) HEARD J.  
PA (JIAN/) JIANG C.  
PA (KEDD/) KEDDIE J.  
PA (ADAM/) ADAM L.  
PA (RATC/) RATCLIFF O.  
PA (REUB/) REUBER J L.  
PA (RIEC/) RIECHMANN J L.  
PA (YUGG/) YU G.  
PA (PINE/) PINEDA O.  
XX  
XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J;  
PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;  
XX WPI; 2002-292022/33.  
DR  
XX P-PSDB; AAU93098.  
DR  
XX An isolated or recombinant polynucleotide used to produce a transgenic  
PT plant.  
XX  
XX Claim 4; Page 558-559; 941pp; English.  
PS  
XX The invention relates to 1 of 232 isolated or recombinant polynucleotides  
CC encoding an Arabidopsis thaliana transcription factor, their variants,  
CC complementary fragments, or related polynucleotide with 318 to 958  
CC sequence identity, where the plant possesses an altered trait as compared  
CC to a wild-type or reference plant, or the plant exhibits an altered  
CC phenotype as compared to a wild-type or reference plant, or the plant  
CC exhibits ectopic expression or altered expression of one or more genes  
CC associated with a plant trait as compared to a wild plant. Also included  
CC are a transgenic plant comprising the polynucleotides, a computer  
CC readable medium having stored sequence information, and identifying a  
CC homologous sequence from a database comprising a plurality of known plant  
CC sequences comprising inputting sequence information selected from one of  
CC 464 fully defined sequences given in the specification. The isolated or  
CC recombinant polynucleotide is used for producing a plant having a  
CC modified trait, the method comprising selecting a polynucleotide that  
CC encodes a polypeptide or an antisense nucleic acid, inserting the  
CC polynucleotide or antisense nucleic acid into an expression vector,  
CC introducing the vector into a plant or a cell of a plant to overexpress  
CC the polypeptide or antisense nucleic acid, thereby producing a modified  
CC plant, and selecting for a modified trait (e.g. increased production of  
CC agriculturally useful proteins or metabolic chemicals, pest tolerance,  
CC environmental stress response (e.g. drought), microbial disease  
CC resistance, herbicide resistance, seed and fruit yield, growth rate, leaf  
CC and flower senescence, and many other traits listed in the specification).  
CC The present sequence is one of the 232 polynucleotides encoding an A.  
CC thaliana transcription factor  
XX  
XX Sequence 969 BP; 267 A; 247 C; 212 G; 243 T; 0 U; 0 Other;  
SQ  
Query Match 5.7%; Score 31.8; DB 6; Length 969;  
QY

Best Local Similarity 54.8%; Pred. No. 5.8;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Dy 89 CTCAACCATTTAGCTACAACACTCCTGCTATGATGATGAGGCATTTCGGCGCCCTTG 148  
D6 392 CTCAAACCAACCAACAACCTCCCTTGTCTCCCTGAAAAGCAGGTGTCACGACGA 451

Qy 149 ACCGATCGCTCACTCATGCTCCCTTAATGCTTTAGTCGACACACTTGGCCGG 203  
Db 452 AACGCCCTCGTCCACTGTCCAGGATAAGATAGAGTCAAAGACAAATGTGTCCGG 506

RESULT 15  
ABZ13985  
ID ABZ13985 standard; DNA; 969 BP.

Db 452 AACGCCCTCGTCCACCTGTCCAGGATAAGATAGAGTCAAAGACAATGTGTCCG 506

Search completed: April 30, 2004, 11:04:24  
Job time : 233.167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:27 ; Search time 1575.49 Seconds  
(without alignments)  
10519.601 Million cell updates/sec

Title: US-10-603-260-4

Perfect score: 555

Sequence: 1 atgcagatattcttctgttca.....cttcacagagaagcgcgtc 555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: EST:\*  
2: em\_estba:\*  
3: em\_estba:\*  
4: em\_estba:\*  
5: em\_estba:\*  
6: em\_estba:\*  
7: em\_estba:\*  
8: em\_estba:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estfun:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_hum:\*  
20: em\_gss\_hum:\*  
21: em\_gss\_hum:\*  
22: em\_gss\_hum:\*  
23: em\_gss\_hum:\*  
24: em\_gss\_hum:\*  
25: em\_gss\_hum:\*  
26: em\_gss\_hum:\*  
27: em\_gss\_hum:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.2	6.5	526	12	BG924185
2	35.6	6.4	583	12	BG776621
3	35.6	6.4	704	10	BF211558
4	35.6	6.4	715	12	BG493794

Result No.	Score	Query Match	Length	ID	Description
5	35.6	6.4	816	12	BG4937402
6	35.2	6.3	498	28	AQ451880
7	35.2	6.3	1201	23	BX381961
8	35	6.3	942	29	CG098039
9	34.6	6.2	931	29	CG711514
10	34.2	6.2	742	14	CA808007
11	33.8	6.1	443	13	BQ494475
12	33.8	6.1	443	13	BQ493352
13	33.8	6.1	485	13	BQ493441
14	33.8	6.1	593	13	BE776398
15	33.8	6.1	624	13	BX311247
16	33.8	6.1	770	14	CF408002
17	33.8	6.1	787	29	CNS01087
18	33.8	6.1	797	29	CC554553
19	33.6	6.1	435	13	BX34834
20	33.6	6.1	485	13	BQ493087
21	33.6	6.1	575	10	AM42644
22	33.6	6.1	606	13	BQ262340
23	33.6	6.1	661	12	BQ302675
24	33.6	6.1	1223	12	B1952284
25	33.4	6.0	1201	13	BX376686
26	33.4	6.0	1456	12	B1757625
27	33.2	6.0	228	9	A1210254
28	33.2	6.0	540	10	AM116871
29	33.2	6.0	628	12	BM332415
30	33.2	6.0	766	28	BZ193724
31	33.2	6.0	765	14	CA809511
32	33.2	6.0	845	14	CA809398
33	33.2	6.0	1068	29	CNS037RT
34	33	5.9	247	9	A1209463
35	33	5.9	288	9	A1209943
36	33	5.9	396	13	BY582873
37	33	5.9	441	28	AQ110347
38	33	5.9	733	28	CC172630
39	33	5.9	977	29	CG684361
40	32.8	5.9	722	28	BH715417
41	32.8	5.9	722	28	BH533401
42	32.8	5.9	1047	29	CNS026FO
43	32.8	5.9	1101	29	CNS00PFO
44	32.6	5.9	495	13	BQ107442
45	32.6	5.9	495	14	CF665213

## ALIGNMENTS

RESULT 1  
LOCUS  
DEFINITION  
BG924185  
526 bp, mRNA, linear, EST 05-JUN-2001  
Oo\_ad\_03B06 LambdaG111FOR Osteragia osteragi adults Osteragia  
Osteragi CDNA clone Oo\_ad\_03B06 5' similar to pIR11628  
propionyl-CoA carboxylase (EC 6.4.1.3) beta chain f528.1, mRNA  
sequence.  
ACCESSION  
BG924185  
VERSION  
BG924185.1 GI:14317857  
SOURCE  
EST  
ORGANISM  
Osteragia osteragi  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongyloidea;  
Trichostrongyloidea; Haemonchidae; Osteragiinae; Osteragia.  
REFERENCE  
1 (bases 1 to 526)  
Blaxter M.L., Parkinson J., Whitton C., Daub J., Guillano D.,  
Hall N., Quayle M. and Barrett B.  
Edinburgh University/Sanger Centre Nematode EST Project  
Unpublished (2000)  
Contact: Blaxter ML  
Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JF, UK.  
Tel: +44 131 650 6760  
Fax: +44 131 670 5450  
Email: mark.blaxter@ed.ac.uk



3' adaptor sequence:  
5'-ATTCTAGAGCCGAGCCGCGACATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size  
1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained  
inserts by PCR. This library was enriched for full-length  
clones and was constructed by Clontech Laboratories (Palo  
Alto, CA)."

## ORIGIN

Query Match 6.4%; Score 35.6; DB 10; Length 704;  
Best Local Similarity 52.7%; Pred. No. 6.6;  
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 ATGCAGATATCTTCTGTCATGACCTATATGACGCTTGTATGATCCGCTAGT 60  
DB 502 ATACACTTCTTCTTCTTGAAGGAATCAATTAAGAGGTTGGTACCAATGCTAGT 443  
QY 61 CATCGTCGATTAATGGGTTATCTGTAACCAACCATTAAGCTACACTGCTATC 120  
DB 442 TCTTATCAAGATAAATACTATTTCATGGAAAAGCCTTGGTCTACTGAATCCCTTC 383  
QY 121 GATGATGAGCCATTTTGGCCGCT 146  
DB 382 CATGTAAAGAGAGTGTGAGCCT 357

## RESULT 4

BG493794/c 715 bp mRNA linear EST 27-MAR-2001

LOCUS 602542056F1 NIH\_MGC\_59 Homo sapiens CDNA clone IMAGE:4673149 5',  
DEFINITION mRNA sequence.

ACCESSION BG493794

VERSION BG493794.1

KEYWORDS GI:13455308

SOURCE EST.

ORGANISM Homo sapiens (human)

REFERENCE Homo sapiens (human)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

EMAIL: cgaabs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.llnl.gov

Plate: LNCM1490 row: p column: 14

High quality sequence stop: 715.

Location/Qualifiers

1. 715

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4673149"

/tissue\_type="mucoepidermoid carcinoma"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_1lb="NIH\_MGC\_59"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site\_1:

Site1 (ggcgcctcgcc); Site\_2: Site1 (ggcctatggc);

Double-stranded CDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGACC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCCGAGCCGCGACATG-dt(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC

## ORIGIN

Library."

Query Match 6.4%; Score 35.6; DB 12; Length 715;  
Best Local Similarity 52.7%; Pred. No. 6.7;  
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 ATGCAGATATCTTCTGTCATGACCTATATGACGCTTGTATGATCCGCTAGT 60  
DB 446 ATACACTTCTTCTTCTTGAAGGAATCAATTAAGAGGTTGGTACCAATGCTAGT 387  
QY 61 CATCGTCGATTAATGGGTTATCTGTAACCAACCATTAAGCTACACTGCTATC 120  
DB 386 TCTTATCAAGATAAATACTATTTCATGGAAAAGCCTTGGTCTACTGAATCCCTTC 327  
QY 121 GATGATGAGCCATTTTGGCCGCT 146  
DB 326 CATGTAAAGAGAGTGTGAGCCT 301

## RESULT 5

BG497402/c 816 bp mRNA linear EST 27-MAR-2001

LOCUS 602538689F1 NIH\_MGC\_59 Homo sapiens CDNA clone IMAGE:4659742 5',  
DEFINITION mRNA sequence.

ACCESSION BG497402

VERSION BG497402.1

KEYWORDS GI:13458919

SOURCE EST.

ORGANISM Homo sapiens (human)

REFERENCE Homo sapiens (human)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

EMAIL: cgaabs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.llnl.gov

Plate: LNCM1456 row: a column: 23

High quality sequence stop: 715.

Location/Qualifiers

1. 816

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4659742"

/tissue\_type="mucoepidermoid carcinoma"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_1lb="NIH\_MGC\_59"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site\_1:

Site1 (ggcgcctcgcc); Site\_2: Site1 (ggcctatggc);

Double-stranded CDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGACC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCCGAGCCGCGACATG-dt(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC

## ORIGIN

Query Match 6.4%; Score 35.6; DB 12; Length 816;  
Best Local Similarity 52.7%; Pred. No. 7.1;  
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy	1	ATGCAATTAATCTTGTTCAAGCACTCAATGATGATGCTGGTAATGCAATCCGCTTAGT	60
Db	448	ATCACTTCTTCCTTTGAGGGAATCAATTTAAGAGAGTTGGTAACCAATGCTAGGT	389
Qy	61	CATCGCTGCATAAATTTGGGTTATGATGCTCAACCAATTGACATCACTCGGATAC	120
Db	388	TCTTATCAAGATTAATACTATTTCATAGGAAAAGCTTTGGTCTATCGAATACCTCTC	329
Qy	121	GATGATGAGCCATTTTTCGCGCCT	146
Db	328	CATGTAAGAAAGAGTTGGTAGCCT	303

RESULT 6	498 bp	DNA	linear	GSS 21-APR-1999
AQ451880				
LOCUS				
DEFINITION				
HQ5184.B1 A05 SP6E RPCT-11 Human Male BAC library Homo sapiens genomic clone Plate=760 Col=9 Row=B, genomic survey sequence.				

ACCESSION	A0451880	
VERSION	A0451880.1	GI:4593050
KEYWORDS	GSS.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Crinata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.  
1 (bases 1 to 498)  
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,

**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
**KEYWORDS** sequence-tagged connectors; mapping; scanning; human genome

MEDLINE 99380589  
PUBMED 10449764  
COMMENT Contact: Mahalras GG, Wallace JC, Hood L

High throughput sequencing center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RCP1-1. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACpac Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end Web Server: <http://www.htsc.washington.edu>  
plate: 760 row: B column: 9  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 498.

```

FEATURES
SOURCE
Location/Qualifiers
1. .498
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=760 Col=9 Row=B"
/sex="male"
/clone_1ib="RPC1-11 Human Male BAC library"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

```

ORIGIN	
Query Match	6.3%; Score 35.2; DB 28; Length 498;
Best Local Similarity	58.7%; Pred. No. 7.6;
Matches	6; Conservative 0; Mismatches 43; Indels 0; Gaps 0
11	TTCTTGTGATGAGACTCTTATGATGATGGCTGTGATGATCCGCTTATGATGATGCTGCG 70

Db 96 TTCGAGGCTGGACCTCTTAGGCCCATGTCATATGATTCATTGGCCATCAATGTTATG 155  
 Qy 71 ATAAATGGGTTATGACATCAACAACATTGTCACACACTACAC 114  
 Db 156 ATAAATATTCTTCACTTCTCTAGAAATGGATGAGTTTACCT 199

RESULT 7	LOCUS	DEFINITION	EST 08-MAY-2003
BX381961	1201 bp	linear	
BX381961	Homo sapiens PLACENTA COT 25-NORMALIZED	Homo sapiens CDNA	
clone CS0D10721F05 3-PRIME, mRNA sequence.			

ACCESSION	EA381961
VERSION	EX381961.1
KEYWORDS	GI:30453007
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens

**REFERENCE**  
**AUTHORS**

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 1201)  
L.J.W.B., Gruber, C., Tesse, J. and Polayes, D.

**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France

Email: [seguie@genoscope.cns.fr](mailto:seguie@genoscope.cns.fr), web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact: Feng Liang Email: [fliang@lifetech.com](mailto:fliang@lifetech.com) URL: <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID: CS0D1072CC03NP1.

```

FEATURES
    source
        location/Qualifiers
            1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /db_xref="CSD10721F05"
                /clone="C80D10721F05"
                /tissue_type="PLACENTA COT 25-NORMALIZED"
                /clone_1ib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                /note="1st strand cDNA was primed with a NotI-oligo(dT)
                    primer. Five prime end enriched, double-strand cDNA was
                    digested with Not I and cloned into the Not I and EcoR V
                    sites of the pCMVSPORT 6 vector. Library was normalized."

```

Query Match 6.3%; Score 35.2; DB 13; Length 1201;  
Best Local Similarity 4.9%; Pred. No. 11;  
Matches 23; Conservative 184; Mismatches 264; Indels 0; Gaps 0;

DQ  
615 CTTGTTATCGAAGCTTATTGCATGGTGTAATGTCACGCCGTACTCATCCTCGCAT 72

Dd  
615 CAHKKKKGCKRHHNNNKNNNNNNNKIKKNNNNKNNNNKNNNNNNNNNNNNNNNN 674

**D**

QY	73	AAATGGGTTACGTACTCAACCATTAGTACACTCGCTAGATGATGAGGCC	132
	:: :: :	: :   : : :	
Db	675	NNNNKKKKKKNNNNNNNNNNNNNNKKKKTNKKGNALGNNINNNNNNNKN	734

[illegible][illegible]

Db 855 MCKMMNNCKAKAMKCDNNNNNNMCMCMCKKKNNMMBGKCMCKMCMCKKKKKHK 914

QY 313 GACCAATTAGTTAGGGGTGGCACTAGGTAACTCAGCAGATTGGGTAAAGAACAC 372

Db 915 NMATKGMCKKBMCKKQKQVNYIKGMVTKKKGAAGYBKCVNKKQKKQCKKQVNGIKYCAKKK 974



QY	373	GACGACGATCCCGGTATCCACAAATAATCAGCAATATTGACAGAAAGATACCTTTAAGG	432
DB	975	KMCMTVKYKMCIDBKMKCKKKKKYCKKCMCMAMMMCMKMYMBAMMMBAHYVM	1034
QY	433	CTGCGACGCTTTTACTGTGCGGATCCACTGTGACTCGATCGATGTGATCCGTACA	483
DB	1035	MMANNMMAMCKMYVMKKKKMKKKKMBMSKMMMMNKVKCMBMBCKSKGCM	1085
RESULT 8			
CG098039		942 bp	DNA
LOCUS			
DEFINITION			
CG098039			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Zea mays			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD			
clade; Panicoideae; Andropogoneae; Zea.			
1 (bases 1 to 942)			
WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,			
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,D. and			
Bennetzen,J.			
Maize Genomic Consortium			
Unpublished (2003)			
CONTACT: Cathy WhiteLaw			
TIGR			
9712 Medical Center Drive, Rockville, MD 20850, USA			
Tel: 301-838-5843			
Fax: 301-838-0208			
Email: whiteLaw@cigr.org			
Seq primer: TP			
Class: sheared ends.			
FEATURES			
source			
1..942			
/organism="Zea mays"			
/mol type="genomic DNA"			
/strain="B73"			
/db xref="taxon:4577"			
/clone="ZMMBTA0693F19"			
/clone_lib="ZM_0.6_1.0 KB"			
/note="vector: PCR4-TOPO; site_1: EcoRI; 0.6-1.0 kb high			
COT selected genomic DNA library"			
ORIGIN			
Query Match			
Best local Similarity			
Matches			
294			
408			
354			
468			
414			
528			
RESULT 9			
CC711514			
LOCUS			
DEFINITION			
CG711514			
VERSION			
KEYWORDS			
CC711514			
GI:32116290			
GSS			
931 bp			
DNA			
linear			
GSS 19-JUN-2003			
OGERR45TC ZM 0.7 1.5 KB zea mays genomic clone ZMMBWA0501G18,			
genomic survey sequence.			
CC711514			
GI:32116290			
GSS			

SOURCE	Zea mays
ORGANISM	Zea mays
REFERENCE	Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, Liliopsida, Poales, Poaceae, PACCAD clade, Panicoideae, Andropogoneae, Zea.
AUTHORS	1 (bases 1 to 931) Whitelaw, C.A., Quackenbush, J., Van Aken, S., Uterback, T., Renick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Clerk, R.W., Nummer, A., Robbins, D. and Lakey, N.
TITLE	Consortium for Maize Genomics
JOURNAL	Unpublished (2002)
COMMENT	Contact: Cathy Whitelaw TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-838-5843 Fax: 301-838-0208 Email: whitelaw@tigr.org Seq primer: 7P Class: sheared ends.
FEATURES	location/Qualifiers 1..931 /organism="Zea mays" /mol_type="genomic DNA" /strain="B73" /db_xref="taxon:4577" /clone="ZMMBMA0501G18" /clone_1lb="ZM 0.7-1.5 KB" /note="vector: pBCSk-; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN	
Query Match	6.2%; Score 34.6; DB 29; Length 931;
Best Local Similarity	47.5%; Pred. No. 16;
Matches 103; Conservative	0; Mismatches 114; Indels 0; Gaps 0;
OY	202 GGATTGGTATCAACGTTATCTAGATCCGCGGCACCCGCTGTGAACCCCTCCCAT 261
DB	96 GGAAGCTTATCTTATCTTCTTCCACGAGATCGAGTGCCTCAAAATTCATTGCG 155
OY	262 GTCGTCGCACATCGGCTCACCTTGCAGAGAGCTTCATGTCAATAAATTTGACATTA 321
DB	156 CGCTCCGCATCATCAGATGAGATTTCTCGACCTTCTATATTAACAAGATTTGTTAAATT 215
OY	322 GGTTTAGGGGTGCACATAGGTAATTCAGCAGAATTTGGTTAAAGAACCAGACGAA 381
DB	216 GATTAAAGGGGTGTAGCAATATCTTATTTGGGTGTCATAGTAAAAAGGGTT 275
OY	382 TCCGCTATCCACAAAATCAGAGCATGATTCAGGAA 418
DB	276 TAGGCCCTCGAAAAAAGAAAGATTAATAAAAAAGAA 312
RESULT 10	
LOCUS	CAB08007 742 bp mRNA linear EST 10-Apr-2003
DEFINITION	CA12110.1IF_H11 Gabernet Sauvignon leaf - CA1211 Vitis vinifera
VERSION	CAB08007
KEYWORDS	CAB08007.1 GI:26256944
SOURCE	EST.
ORGANISM	Vitis vinifera
REFERENCE	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
AUTHORS	1 (bases 1 to 742) Goes da Silva, F., Lim, H., Iandolo, A., Baek, J., Jones, K., Walker, M.A. and Cook, D.R.
TITLE	Transcriptional responses of Vitis vinifera to infection by the bacterial pathogen Xylella fastidiosa
JOURNAL	Unpublished (2003)
COMMENT	Contact: Doug Cook CABs Genome Facility

UC Davis Department of Plant Pathology  
1 Shields Ave., Davis, CA 95616, USA  
Tel: 530 754 6561  
Fax: 530 754 6617  
Email: drcoc@ucdavis.edu  
Seq primer: GTTATCAGTCACGGTACC.  
Location/Qualifiers

## FEATURES

source

1..742

/organism="Vitis vinifera"

/mol\_type="mRNA"

/cultivar="Cabernet Sauvignon"

/db\_xref="taxon:29760"

/clone="CA1211041F.H11"

/sex="hermaphrodite"

/dev\_stage="late season sample"

/lab\_host="D55alpha"

/clone\_1ib="Cabernet Sauvignon leaf - CA1211"

/note="Organ: Leaf; Vector: pDMR; Site 1: Sfil; Site 2: Sfil; CA1211 is a cDNA library of Cabernet Sauvignon leaves. The leaves were collected on September 20, 2001, in Napa Valley, California, and represent leaves in late season development. These leaves were symptomatic and verified to be infected with the bacterial pathogen, Xylella fastidiosa, based on a diagnostic assay using PCR and Xylella-specific primer pairs. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:

5'-AAGCAGTGTATCAGCAGATGCGCTTACGCGCGG-3' and  
5'-ATTCTAGAGCCGAGCGCGGCGCAGATG-3' (30'NN-3'). Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

## ORIGIN

Query Match 6.2%; Score 34.2; DB 14; Length 742;  
Best Local Similarity 56.8%; Pred. No. 19;  
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 336 ACTAGTATTCAGACAGATTTGGTTAAAGAACACAGCAGCATCCGCTATCCACA 395  
DB 616 ATTAGCTTATTAGTTGTTATTGTTATTAAGAACTGACACTTAAATCTTATCCGAG 675  
QY 396 AAAATCAGGCGATATTCAGACAGCATCTCTTACGCGCTGCGACGCTTTT 446  
DB 676 CAATCAGGCGCATTTGTTAAAGATGTAGCTCAAGAGGTTCTCTCTTTT 726

## RESULT 11

BQ494475/c

LOCUS BQ494475 434 bp mRNA linear EST 31-OCT-2002  
DEFINITION EST03641 Paracoccidioides brasiliensis cDNA, mRNA sequence.  
ACCESSION BQ494475  
VERSION BQ494475.1 GI:24443116  
KEYWORDS EST.

## SOURCE

ORGANISM

Paracoccidioides brasiliensis  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Omygenales; mitosporic Omygenales; Paracoccidioides.

1 (bases 1 to 434)  
Goldman, G.H., dos Reis Marques, E., Duarte Ribeiro, D.C., de Souza  
Bernardes, L.A., Quiapin, A.C., Vitorrelli, P.M., Savoldi, M.,  
Semghini, C.P., de Oliveira, R.C., Nunes, L.R., Travassos, L.R.,  
Puccia, R., Batista, W.L., Ferreira, L.E., Moreira, J.C.,  
Bogossian, A.P., Tekala, F., Nobrega, M.P., Nobrega, F.G. and  
Goldman, M.H.

## REFERENCE

AUTHORS

## TITLE

Expressed sequence tag analysis of the human pathogen  
Paracoccidioides brasiliensis yeast phase: identification of  
putative homologues of Candida albicans virulence and pathogenicity  
genes

## JOURNAL

COMMENT

Contact: Gustavo Henrique Goldman  
Laboratory of Molecular Biology  
Universidade de Sao Paulo - USP - FCFRP  
Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil

Email: ggoldman@usp.br.  
Location/Qualifiers  
1..434  
/organism="Paracoccidioides brasiliensis"  
/mol\_type="mRNA"  
/db\_xref="taxon:121759"  
/clone\_1ib="PB0001"

## ORIGIN

Query Match 6.1%; Score 33.8; DB 13; Length 434;  
Best Local Similarity 54.4%; Pred. No. 20;  
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 429 AGGCGTGGCGACGCTTTTACTGCGGATCCACTGACCTCCAGTGGTACGTCAGTGA 488  
DB 352 AGGCGTGGCTACCCAGTACGATGTCGCGAGGAGATTACGAGAGTACCTCTGACAA 293  
QY 489 AGAAGCAAAATAGCTGCGATGACATCATATCGCATACACCACTTCATACAGAA 548  
DB 292 AGAAGACGCGCAGATACCGGCGTAGATGAACCCCAACCAATCACCCTATAC 233  
QY 549 TGCTG 553  
DB 232 CGCTG 228

## RESULT 12

BQ493352/c

LOCUS BQ493352 443 bp mRNA linear EST 31-OCT-2002  
DEFINITION EST02518 PB0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.  
ACCESSION BQ493352  
VERSION BQ493352.1 GI:24440281  
KEYWORDS EST.

## SOURCE

ORGANISM

Paracoccidioides brasiliensis  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Omygenales; mitosporic Omygenales; Paracoccidioides.

1 (bases 1 to 443)  
Goldman, G.H., dos Reis Marques, E., Duarte Ribeiro, D.C., de Souza  
Bernardes, L.A., Quiapin, A.C., Vitorrelli, P.M., Savoldi, M.,  
Semghini, C.P., de Oliveira, R.C., Nunes, L.R., Travassos, L.R.,  
Puccia, R., Batista, W.L., Ferreira, L.E., Moreira, J.C.,  
Bogossian, A.P., Tekala, F., Nobrega, M.P., Nobrega, F.G. and  
Goldman, M.H.

## REFERENCE

AUTHORS

## TITLE

Expressed sequence tag analysis of the human pathogen  
Paracoccidioides brasiliensis yeast phase: identification of  
putative homologues of Candida albicans virulence and pathogenicity  
genes

Contact: Gustavo Henrique Goldman  
Laboratory of Molecular Biology  
Universidade de Sao Paulo - USP - FCFRP  
Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil  
Email: ggoldman@usp.br.

## JOURNAL

COMMENT

## FEATURES

source

1..443  
/organism="Paracoccidioides brasiliensis"  
/mol\_type="mRNA"  
/db\_xref="taxon:121759"  
/clone\_1ib="PB0001"

## ORIGIN

source

Query Match 6.1%; Score 33.8; DB 13; Length 443;  
Best Local Similarity 54.4%; Pred. No. 20;  
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 429 AGGCGTGGCGACGCTTTTACTGCGGATCCACTGACCTCCAGTGGTACGTCAGTGA 488  
DB 188 AGGCGTGGCTACCCAGTACGATGTCGCGAGGAGTTGCGAGAGTACCTCTGACAA 129  
QY 489 AGAAGCAAAATAGCTGCGATGACATCATATCGCATACACCACTTCATACAGAA 548  
DB 128 AGAAGACGCGCAGATACCGGCGTAGATGAACCCCAACCAATCACCCTATAC 69

Oy 549 TGCTG 553  
 Db 68 CGCTG 64  
 RESULT 13  
 BQ499441/c 485 bp mRNA linear EST 31-OCT-2002  
 LOCUS BQ499441/c  
 DEFINITION EST08666 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.  
 ACCESSION BQ499441  
 VERSION BQ499441.1 GI:24452737  
 KEYWORDS EST.  
 SOURCE Paracoccidioides brasiliensis  
 ORGANISM Paracoccidioides brasiliensis  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Onygenales; microsporid Onygenales; Paracoccidioides.  
 REFERENCE 1 (bases 1 to 485)  
 Goldman, G.H., dos Reis Marques, E., Duarte Ribeiro, D.C., de Souza  
 Bernardes, L.A., Quispin, A.C., Vitorelli, P.M., Savoldi, M.,  
 Semighini, C.P., de Oliveira, R.C., Nunes, L.R., Travaes, L.R.,  
 Puccia, R., Batista, M.L., Ferreira, L.E., Moreira, U.C.,  
 Bogosian, A.P., Tekala, F., Nobrega, M.P., Nobrega, F.G. and  
 Goldman, M.H.  
 Expressed sequence tag analysis of the human pathogen  
 Paracoccidioides brasiliensis yeast phase: identification of  
 putative homologues of Candida albicans virulence and pathogenicity  
 genes  
 TITLE Eukaryot. Cell 2 (1), 34-48 (2003)  
 JOURNAL Contact: Gustavo Henrique Goldman  
 COMMENT Laboratory of Molecular Biology  
 Universidade de Sao Paulo - USP - FCBP  
 Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil  
 Email: ggoldman@usp.br  
 Location/Qualifiers  
 source  
 1. 485  
 /organism="Paracoccidioides brasiliensis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:121759"  
 /clone\_lib="Pb0001"  
 ORIGIN  
 Query Match 6.1%; Score 33.8; DB 13; Length 485;  
 Best Local Similarity 54.4%; Pred. No. 21;  
 Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
 Oy 429 AGGGCTGGCGACCTTTTACTGCGGATCCATCGATCCGATCGTACCGTCAACAGTACA 488  
 Db 239 AGGGCTGGCTCAACCCAGGAGTAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180  
 Oy 489 AGAACAACAAATATAGCTGGCATGACGATCATATCGCGATATCCACCACTTCATACGAGAA 548  
 Db 179 AGAAGACCGCGAGATACCGGCTGATGAAAGCCCAACAGTACCCCAACCTTATAC 120  
 Oy 549 TGCTG 553  
 Db 119 CGCTG 115  
 RESULT 14  
 BE776398 593 bp mRNA linear EST 20-SEP-2000  
 LOCUS BE776398  
 DEFINITION MY-15-C-12 PinfeestansMY Phytophthora infestans cDNA, mRNA sequence.  
 ACCESSION BE776398  
 VERSION BE776398.1 GI:10230053  
 KEYWORDS EST.  
 SOURCE Phytophthora infestans (potato late blight agent)  
 ORGANISM Phytophthora infestans  
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
 Phytophthora.  
 REFERENCE 1 (bases 1 to 593)  
 Kamoun, S., Hirber, P.T., Sobral, B.W.S., Nuss, D. and Govers, F.  
 Initial assessment of gene diversity for the oomycete pathogen

JOURNAL Phytophthora infestans based on expressed sequences  
 FUNGAL Genet. Biol. 28 (2), 94-106 (1999)  
 MEDLINE 20056376  
 PUBMED 10587472  
 COMMENT Contact: Govers F  
 Laboratory of Phytopathology  
 Wageningen University  
 Bomenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands  
 Tel: 31 317 483 138  
 Fax: 31 317 483 412  
 Email: Francine.Govers@medew.fyto.wau.nl.  
 Location/Qualifiers  
 source  
 1. 593  
 /organism="Phytophthora infestans"  
 /mol\_type="mRNA"  
 /strain="DDR7602, A1 mating type"  
 /db\_xref="taxon:4787"  
 /dev\_stage="4-week old vegetative, non-sporulating  
 mycelium in synthetic medium"  
 /lab\_host="E. coli, strain DH5-alpha"  
 /clone\_lib="PinfeestansMY"  
 /note="Vector: pSPORT1, Site 1: SalI, Site 2: NotI; Total  
 RNA was isolated from mycelium of P. infestans DDR7602  
 cultured for 4 weeks in synthetic medium. EST clones were  
 named by their position in the microtiter plate, preceded  
 by the prefix MY (for mycelial) and the successive number  
 of the microtiter plate (e.g. MY-06-A-04)."  
 ORIGIN  
 Query Match 6.1%; Score 33.8; DB 10; Length 593;  
 Best Local Similarity 49.7%; Pred. No. 23;  
 Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
 Oy 349 GCAGATTGGGTAAAGACACGACGACGATCCCGCTATCCAAATATGAGGCACT 408  
 Db 110 GCTGAATTAGCCACTTAATATCCGCAATGGCAGCGAAAGTCTCAAGTTGGCGGC 169  
 Oy 409 ATTGCAGAAACGATACCTTTAGGCGCTCGCAGACCTTTTACTCGCGCATCGACTGACTCC 468  
 Db 170 ACATCCGTGGGAGGAGTCCGCGAGACTCTGCGGCTGTGACGATCGTCCGAGGAGAGA 229  
 Oy 469 GATGATACCGTCACTAGATGAAACAAATATGCTGGCATGACATATAT 521  
 Db 230 GATCGGCTGTGGCCGTGATGTCCGCAATGGGCGCACACGACGATATTT 282  
 RESULT 15  
 BX311247/c 624 bp mRNA linear EST 19-DEC-2003  
 LOCUS BX311247/c  
 DEFINITION BX311247 AGENNE Rainbow trout multi-tissues subtracted library  
 (tcay) Oncorhynchus mykiss cDNA clone tcay0023b.h.08 3prim, mRNA  
 sequence.  
 ACCESSION BX311247  
 VERSION BX311247.2 GI:40226031  
 KEYWORDS EST.  
 SOURCE Oncorhynchus mykiss (rainbow trout)  
 ORGANISM Oncorhynchus mykiss  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 REFERENCE 1 (bases 1 to 624)  
 Govoroun, M., Guilgen, Y. and Le Gac, F.  
 Construction and primary characterization of normalized cDNA  
 libraries in rainbow trout, Oncorhynchus mykiss  
 Unpublished (2003)  
 On Apr 7, 2003 this sequence version replaced GI:29591892.  
 COMMENT Contact: Guilgen Y  
 INRA - SCRIBE  
 Campus de Beaulieu, RENNES cedex, 35042, France  
 Tel: 02.23.48.50.09  
 Fax: 02.23.48.50.20  
 Email: Yann.Guilgen@beaulieu.rennes.inra.fr  
 Sequence cleaned of vector, adaptor and repetitions. Contact us



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 09:53:32 ; Search time 65.2277 Seconds  
(without alignments)  
4721.887 Million cell updates/sec

Title: US-10-603-260-4

Perfect score: 555

Sequence: 1 atgcagatattcttctgttca.....cttcatacagagatcgtcgtc 555

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34.6	6.2	7218	1 US-08-232-463-14	Sequence 14, Appl
2	34.4	6.2	2277	1 US-08-676-967-2	Sequence 2, Appl
3	34.4	6.2	2277	1 US-08-676-974-2	Sequence 2, Appl
4	34.4	6.2	2277	2 US-08-098-487-2	Sequence 203, Appl
5	31.8	5.7	4715	4 US-08-956-171E-203	Sequence 2564, Ap
6	31.6	5.7	2217	4 US-09-107-532A-2564	Sequence 162, App
7	31.6	5.6	918	2 US-08-743-637B-162	Sequence 162, App
8	31.6	5.6	918	3 US-08-526-840B-162	Sequence 162, App
9	31.6	5.6	5162	3 US-09-298-367B-1	Sequence 1, Appl
10	31.6	5.6	5162	3 US-09-298-367B-4	Sequence 1, Appl
11	31.6	5.6	5262	3 US-09-298-367B-5	Sequence 5, Appl
12	31.6	5.6	6850	3 US-09-298-367B-2	Sequence 2, Appl
13	30.8	5.5	33248	4 US-09-596-002-24	Sequence 24, Appl
14	30.6	5.5	1488	4 US-09-543-681A-2188	Sequence 2188, Ap
15	30.4	5.5	1722	4 US-09-328-352-2832	Sequence 2832, Ap
16	30.4	5.4	805	4 US-09-540-824-16	Sequence 16, Appl
17	30.4	5.4	1830121	4 US-09-557-884-1	Sequence 1, Appl
18	30.4	5.4	1830121	4 US-09-643-990A-1	Sequence 18, Appl
19	29.8	5.4	2869	4 US-08-809-955-18	Sequence 15, Appl
20	29.8	5.4	1664976	4 US-08-916-421B-1	Sequence 15, Appl
21	29.6	5.3	831	4 US-08-935-263-15	Sequence 15, Appl
22	29.6	5.3	831	4 US-09-594-185-15	Sequence 15, Appl
23	29.6	5.3	1443	4 US-09-107-532A-2640	Sequence 3640, Ap
24	29.6	5.3	6115	2 US-08-808-793-2	Sequence 2, Appl
25	29.6	5.3	6115	2 US-08-772-512A-2	Sequence 1, Appl
26	29.6	5.3	6318	2 US-08-808-793-1	Sequence 1, Appl
27	29.6	5.3	6318	3 US-08-772-512A-1	Sequence 1, Appl

28	29.4	5.3	1239	4 US-09-489-039A-6730	Sequence 6730, Ap
29	29.2	5.3	879	4 US-09-134-000C-989	Sequence 889, App
30	29.2	5.3	1278	4 US-09-491-577-45	Sequence 45, Appl
31	29.2	5.3	15213	4 US-08-961-527-26	Sequence 26, Appl
32	29	5.2	3018	4 US-09-328-352-751	Sequence 751, App
33	29	5.2	17310	4 US-08-956-171E-23	Sequence 23, Appl
34	28.8	5.2	936	3 US-08-910-501-3	Sequence 3, Appl
35	28.8	5.2	936	3 US-09-398-550-3	Sequence 3, Appl
36	28.8	5.2	939	3 US-08-910-501-1	Sequence 1, Appl
37	28.8	5.2	939	3 US-09-398-550-1	Sequence 1, Appl
38	28.8	5.2	2060	4 US-09-023-655-1153	Sequence 1153, Ap
39	28.8	5.2	41100	4 US-09-755-665-46	Sequence 46, Appl
40	28.6	5.2	2623	4 US-09-023-655-307	Sequence 307, App
41	28.6	5.2	3133	1 US-08-162-809-1	Sequence 1, Appl
42	28.6	5.2	4860	4 US-09-328-352-3221	Sequence 3221, Ap
43	28.6	5.2	6281	4 US-09-376-594-1039	Sequence 1039, Ap
44	28.6	5.2	193303	4 US-09-497-855A-37	Sequence 37, Appl
45	28.6	5.2	193303	4 US-09-497-855A-44	Sequence 44, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diegonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMUT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZspc-F18  
US-08-232-463-14  
Query Match 6.2%; Score 34.6; DB 1; Length 7218;



```

GENERAL INFORMATION:
APPLICANT: COLLINS, Kathleen
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent'n Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,487
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: USB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-098-487-2

Query Match      6.2%  Score 34.4;  DB 2;  Length 2277;
Best Local Similarity 34.5%;  Pred. No. 0.065;
Matches 51;  Conservative 25;  Mismatches 72;  Indels 0;  Gaps 0;

QY      11  TTTCTGTTTCAGACGACTCTATGATGAGCTTGATGATCGCGCTAGTCAGTCGCTGC 70
Db      662  TTTTCTTTNACNSWTCCTGTCGTGTTNSWTCRGNSWTTTTCYTNCDDATNGNSWN 603
QY      71  ATAAATGGGTTATCGTACTCAAAACCATTAAGCTAACACTGCTAGTATGATGATGAGG 130
Db      602  ACNSWYTGNGTCTCTATTTATATTTCTCTTNGCNAAGCCGATGCAAGCATGCTGCTCN 543
QY      131  CCATTTTCCGGCGGCTTGACCGATCGCT 158
Db      542  CCYTDAATTCCTTCATRTTATNCCT 515

RESULT 5
US-08-956-171E-203/C
Sequence 203, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gill H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ. ID NO: 203:
SEQUENCE CHARACTERISTICS:
LENGTH: 4715 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 203:
US-08-956-171E-203

Query Match      5.7%; Score 31.6; DB 4; Length 4715;
Best Local Similarity 59.3%; Pred. No. 0.97;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Dy      2  TGCAGATTATCTGTGCATGAGACTCATATAGATGAGCTGGTATGATCCGTTAGTC 61
Db      4275 TTCACTTCATCTTCTTCAATGACTATATCCGATATGATGATTCTTCTTCAACTGCTC 4216
Dy      62  ATCGCTGCATTAATTGGGTATGCTACTGA 92
Db      4215 ATCTTCTGCATCATTTAGTCTTCTTGATCA 4185

RESULT 6
US-09-107-532A-2564
Sequence 2564, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:

```



NAME: Atinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 2564:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2217 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...2217  
SEQUENCE DESCRIPTION: SEQ ID NO: 2564:  
US-09-107-532h-2564  
Query Match  
Best Local Similarity 5.7%; Score 31.6; DB 4; Length 2217;  
Matches 61; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 2 TCGAGATATTTCTTGTTCATGACCTATATGATGCGCTGTGTAATGATCGCTTATGC 61  
DB 1055 TGGAAACGCTATTGTAAGTATTATTTAGAAATGAGTAAAGGCGAAGCTAATGC 1114  
QY 62 ATGCTCGATTAATATGGTATTCGTACTCAAAACCATTCAGTCAACTCA 111  
DB 1115 AGCGTTCGACGATTTGGTTATAGTTCCTCAAAATGACTGTAATATCA 1164  
RESULT 7  
US-08-743-637B-162  
Sequence 162, Application US/08743637B  
Patent No. 5994066  
GENERAL INFORMATION:  
APPLICANT: BERGERON, Michel G.  
APPLICANT: PICARD, Francois J.  
APPLICANT: OUELLETTE, Marc  
APPLICANT: ROY, Paul H.  
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: QUARLES & BRADY  
STREET: 411 EAST WISCONSIN AVENUE  
CITY: MILWAUKEE  
STATE: WISCONSIN  
COUNTRY: USA  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/743,637B  
FILING DATE: 04-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/526,840  
FILING DATE: 11-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 850586.90012

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414) 277-5591  
INFORMATION FOR SEQ ID NO: 162:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 918 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-743-637B-162  
Query Match  
Best Local Similarity 5.6%; Score 31; DB 2; Length 918;  
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
QY 310 ATTGAGCAATTAGTTAGGTTGGGGTGCGACTGTAATTCAGCAGAAATTTGGTTAAAGAA 369  
DB 624 ATTGGCGACACATTAACCGAATCGCAAAAACGATTTGTGAAATGTTGGACATTA 683  
QY 370 CAGACGACGAATCCGCTATCCAAATAATCAGGAGTATTCAGGAACGATACCTTTA 429  
DB 684 CGCAACAGGCAATCATTGATTCGCGCTGCTACGCAACATCGTGAAGTATCAATA 743  
QY 430 GGGCTGGCAGCCCTTTTACTGCGCGATCCACTGACTCGATGTAACCTGACAGTAGAA 489  
DB 744 AAGCGGGCGGGTAATATGATGTGACGCAATGATATTCGGGTTCGATACCAATCG 803  
QY 490 GAAACCAATATAGCTGGCA 508  
DB 804 CAACCGATTGTATGGCA 822  
RESULT 8  
US-08-526-840B-162  
Sequence 162, Application US/08526840B  
Patent No. 6001564  
GENERAL INFORMATION:  
APPLICANT: BERGERON, Michel G.  
APPLICANT: OUELLETTE, Marc  
APPLICANT: ROY, Paul H.  
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND  
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY  
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES  
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...  
NUMBER OF SEQUENCES: 177  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: QUARLES & BRADY  
STREET: 411 East Wisconsin Avenue  
CITY: Milwaukee  
STATE: Wisconsin  
COUNTRY: USA  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/526,840B  
FILING DATE: 11-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/304,732  
FILING DATE: 12-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 850586.90012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414) 277-5591  
INFORMATION FOR SEQ ID NO: 162:

SEQUENCE CHARACTERISTICS:  
LENGTH: 918 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-526-840B-162

Query Match  
Best Local Similarity 47.2%; Score 31; DB 3; Length 918;  
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

310 ATTGACCAATTAGGTTAGGGGCTGACCTAGTAATTGACGAGAAATTGGTTAAAGAA 369  
624 ATTGGGCAACACATTAACCGAATGCCAAAAACGATTTGGGAAATTGGTGGACATTA 683  
370 CACGACGAGAAATCCCGCTATCCAAAAATCAGCAGTATTGACAGAAACGATACCTTTA 429  
684 CGCAACAGGCAATCATGATTCGCGCTGCTACGCCAACATCGTGAAGGTGACATTA 743  
430 GGGCTGCCAGCCTTTTACTGGCGATCCACTGACCTCCGATGGTACCGTACAGTAGAA 489  
744 AAGCGGGCGGGTAATATGGTATGCGAATGATATTGGGGTGGTGCATACCAAAATCG 803  
490 GAAACCAAAATAGCTGGCA 508  
804 CAACCGATTGTGATGGCA 822

RESULT 9  
US-09-298-367B-1/c  
Sequence 1, Application US/09298367B  
Patent No. 6180112

GENERAL INFORMATION:  
APPLICANT: Highlander, Sarah K.  
APPLICANT: Federova, Natalie D.  
TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA VACCINE  
FILE REFERENCE: BCM-03728  
CURRENT APPLICATION NUMBER: US/09/298,367B  
CURRENT FILING DATE: 1999-04-22  
PRIOR APPLICATION NUMBER: 08/834,455  
PRIOR FILING DATE: 1997-04-15  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 5162  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-298-367B-1

Query Match  
Best Local Similarity 47.2%; Score 31; DB 3; Length 5162;  
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

310 ATTGACCAATTAGGTTAGGGGCTGACCTAGTAATTGACGAGAAATTGGTTAAAGAA 369  
3719 ATTGGGCAACACATTAACCGAATGCCAAAAACGATTTGGGAAATTGGTGGACATTA 3660  
370 CACGACGAGAAATCCCGCTATCCAAAAATCAGCAGTATTGACAGAAACGATACCTTTA 429  
3659 CGCAACAGGCAATCATGATTCGCGCTGCTACGCCAACATCGTGAAGGTGACATTA 3660  
430 GGGCTGCCAGCCTTTTACTGGCGATCCACTGACCTCCGATGGTACCGTACAGTAGAA 489  
3599 AAGCGGGCGGGTAATATGGTATGCGAATGATATTGGGGTGGTGCATACCAAAATCG 3540  
490 GAAACCAAAATAGCTGGCA 508  
3539 CAACCGATTGTGATGGCA 3521

RESULT 10  
US-09-298-367B-4/c  
Sequence 4, Application US/09298367B  
Patent No. 6180112

GENERAL INFORMATION:  
APPLICANT: Highlander, Sarah K.  
APPLICANT: Federova, Natalie D.  
TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA VACCINE  
FILE REFERENCE: BCM-03728  
CURRENT APPLICATION NUMBER: US/09/298,367B  
CURRENT FILING DATE: 1999-04-22  
PRIOR APPLICATION NUMBER: 08/834,455  
PRIOR FILING DATE: 1997-04-15  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 4  
LENGTH: 5162  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-298-367B-4

Query Match  
Best Local Similarity 47.2%; Score 31; DB 3; Length 5162;  
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

310 ATTGACCAATTAGGTTAGGGGCTGACCTAGTAATTGACGAGAAATTGGTTAAAGAA 369  
3719 ATTGGGCAACACATTAACCGAATGCCAAAAACGATTTGGGAAATTGGTGGACATTA 3660  
370 CACGACGAGAAATCCCGCTATCCAAAAATCAGCAGTATTGACAGAAACGATACCTTTA 429  
3659 CGCAACAGGCAATCATGATTCGCGCTGCTACGCCAACATCGTGAAGGTGACATTA 3660  
430 GGGCTGCCAGCCTTTTACTGGCGATCCACTGACCTCCGATGGTACCGTACAGTAGAA 489  
3599 AAGCGGGCGGGTAATATGGTATGCGAATGATATTGGGGTGGTGCATACCAAAATCG 3540  
490 GAAACCAAAATAGCTGGCA 508  
3539 CAACCGATTGTGATGGCA 3521

RESULT 11  
US-09-298-367B-5/c  
Sequence 5, Application US/09298367B  
Patent No. 6180112

GENERAL INFORMATION:  
APPLICANT: Highlander, Sarah K.  
APPLICANT: Federova, Natalie D.  
TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA VACCINE  
FILE REFERENCE: BCM-03728  
CURRENT APPLICATION NUMBER: US/09/298,367B  
CURRENT FILING DATE: 1999-04-22  
PRIOR APPLICATION NUMBER: 08/834,455  
PRIOR FILING DATE: 1997-04-15  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 5  
LENGTH: 5262  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-298-367B-5

Query Match  
Best Local Similarity 47.2%; Score 31; DB 3; Length 5262;  
Matches 94; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

310 ATTGACCAATTAGGTTAGGGGCTGACCTAGTAATTGACGAGAAATTGGTTAAAGAA 369

Db	4158	ATTGGGCACACATTAAACCGAATGCAAAAAACAATTTTGTGAAATGGTGGACAATA	4092
Qy	370	CACGACGACGAATCCCGCTATCCACAAAAATCAGGCGATATTCGAGGAACGATACCTTTA	429
Db	4098	CGCAACAGGCAATCCATTGATTTCCGGCTGCTACCCCAACATCGTGAAAAGTGTACCATTA	4038
Qy	430	GGGCTCGGAGAGCCTTTTCTGGCGCATCCACTGACACTCCGATGGTACCGTCAAGTAGAA	489
Db	4038	AAGCGGGGGCGGGTAATATGTGTGTACGCAATGATATTTGCAGTTCGATCCAAATCG	3978
Qy	490	GAAACCAAAATAGCTGGCA	508
Db	3978	CAAAACGATTTGTGATGGCA	3960

```

RESULT 12
US-09-298-367B-2/c
Sequence 2, Application US/09298367B
Patent No. 6180112
GENERAL INFORMATION:
APPLICANT: Highlander, Sarah K.
APPLICANT: Federova, Natalie D.
TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA VACCINE
FILE REFERENCE: BCM-03728
CURRENT APPLICATION NUMBER: US/09/298,367B
CURRENT FILING DATE: 1999-04-22
PRIORITY APPLICATION NUMBER: 08/634,455
PRIORITY FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentn Ver. 2.0
SEQ ID NO 2
LENGTH: 6850
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-298-367B-2

```

	Query Match	Similarity	5 - 6%;	Score 31;	DB 3;	Length 6850;
	Best Local	Similarity	47.2%;	Pred. No. 2.4;		
	Matches	94;	Conservative	0;	Mismatches 105;	Indels 0;
					Gaps	0;
QY	310	ATTGACCAATTAGTTTAGGGGTGGCACTAGATTACAGACAAATTTCGGTTAAAGAA	369			
Db	5407	ATTGGGCAACACATTAAACCGAATCGCAAAAACAGATTTCGTGAATTGGTGGACATTA	5348			
QY	370	CACGACGACGAATCCCGCTATCCAAATAACAGCAGTTTCGACGAAACGATACCTTTA	429			
Db	5347	CGCAACGGGCATTCATTGATTCGGGCTGCTACCCAAACATCTGGAAATGTGCATTA	5288			
QY	430	GGGCTTCGACGAGCTTTTACTTCGCGCATCCACTGGACTCCGATGTACCGTCAACATGAA	489			
Db	5287	AAACGGGGGCGGTAATAATGTGTACGACGAATGATTTGCGGTTCGATACCAATCG	5228			
QY	490	GAACCAAAATAGCTGGCA	508			
Db	5227	CAAAACGATTGTGATGGCA	5209			

RESULT 13  
 US-09-596-002-24/C  
 Sequence 24, Application US/03596002  
 Patent No. 6632636  
 GENERAL INFORMATION:  
 APPLICANT: Lagace, Robert, E.  
 APPLICANT: Patterson, Chandra  
 APPLICANT: Berg, Kim, L.  
 TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME  
 FILE REFERENCE: PM-0008-4 US  
 CURRENT APPLICATION NUMBER: US/09/596,002  
 CURRENT FILING DATE: 2000-06-16  
 PRIOR APPLICATION NUMBER: 60/140,121  
 PRIOR FILING DATE: 1999-06-18

```

; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 33248
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte template ID No. 6632636 24
; PUBLICATION INFORMATION:
US-03-596-002-24

```

Query Match	Similarity	5.5%;	Score 30.8;	DB 4;	Length 33248;
Best Local	Similarity	45.2%;	Pred. No. 8.4;		
Matches	113;	Conservative	0;	Mismatches	137;
				Indels	0;
				Gaps	0;
QY	138	TCGCCGCTTGACCGATCGCTCACTCATGCTCGCCTTAATGCTTAACTGAGACACAGATT	197		
Db	11987	TCATGACACGCCCCCAACGACATTTAAACGCCAGCTTACCGCTTGCTGTGCGCATTTGCTT	11928		
QY	198	GAGCGGATTTGTATCATCAACGTTATCTAAATGCGCGCACCGCTCGCTGGAACCCCTTC	257		
Db	11927	TGGCAAGGGGGCAATATCATCTGTGATTTTAAACACACGCTCAGATTAACCAAAACC	11866		
QY	258	CCATGTCGTGCGCATGCGCTCACTTTGCCAAGAGCTTCATTTGCAATAAAATTGAGCA	317		
Db	11867	GCTTTATTTTGGCAATCAGCTTAATCTCGGTGCGTGGATGATTTTTCAGGGTGGATTT	11808		
QY	318	ATTAGCTTTAGGGGTGGCACTAGATTAATCAGACAAATTTGGGTTAAAGAACACAGACA	377		
Db	11807	AAATTTGCTCACTGTCGCCCTTAATTTGTCAGAGTGAATTTGCTTTAACACAAAGGACGA	11748		
QY	378	CGAATCCCGC	387		
Db	11747	GAATTTATGC	11738		

```

RESULT 14
US-09-543-681A-2188
; Sequence 2188, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2188
; LENGTH: 1488
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2188

```

	Query Match	Similarity	5.5%	Score	No. 6;	DB 4;	Length	1488;
	Best Local	Similarity	49.7%	Pred.	No. 1.2;			
	Matches	78;	Conservative	0;	Mismatches	79;	Indels	0;
							Gaps	0;
QY	209	TGATCAACGTTATCTAGTAATCGCGCCACCGTCTGTGAACCCCTCTCCATGTCGTG	268					
DB	1043	TGACCAAAAAAGCATTTGATTTGCTGAAAACGATGAAAAATGGTTTCTTCTTCAAGTCG	1102					
QY	269	CCATCGGCTACCTTTGCAGAGCCTTCATTGCATTAATAATTGAGCAATTAGCTTTAG	328					
DB	1103	AAACGCGCTCCATTGATTAACAAAGATCATATATGCTTAATCCTTGTGTCCTCAATGGTGA	1162					
QY	329	GGGTGGCACTGTAATTGACGACAAATTTGGGTTAAA	365					
DB	1163	CGTGGCAATTGATGAAAGCGGTCAAAATTTGGTTAGA	1199					



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2004, 15:29:25 ; Search time 246.227 Seconds  
(without alignments)  
10195.950 Million cell updates/sec

Title: US-10-603-260-4

Perfect score: 555

Sequence: 1 atgcagattatctgtcttca.....cttcacagagatctgtt 555

Scoring table:

IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*

13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*

14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*

15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*

17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*

18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	555	100.0	555	US-10-603-260-4	Sequence 4, Appl1
2	555	100.0	2578	US-10-603-260-1	Sequence 1, Appl1
3	36.8	6.6	26997	US-10-221-714A-473	Sequence 473, App
4	33.2	6.0	32185	US-10-087-192-1135	Sequence 1135, App
5	33	5.9	435	US-09-783-590-1460	Sequence 1460, App
6	33	5.9	96499	US-10-087-192-2011	Sequence 2011, App
7	32.8	5.9	948	US-10-282-122A-15368	Sequence 15368, A
8	32.4	5.8	1691139	US-10-067-514-1	Sequence 1, Appl1
9	32.4	5.8	1691139	US-10-419-723-1	Sequence 1, Appl1
10	32.2	5.8	653122	US-10-087-192-226	Sequence 226, App
11	32	5.8	549	US-10-282-122A-18072	Sequence 18072, A
12	31.8	5.7	969	US-09-938-842A-1790	Sequence 1790, App
13	31.8	5.7	969	US-09-934-455-271	Sequence 271, App
14	31.8	5.7	969	US-09-938-842A-1790	Sequence 1790, App

15	31.8	5.7	969	13	US-10-225-066A-377	Sequence 377, App
16	31.8	5.7	969	16	US-10-374-780A-2733	Sequence 2733, App
17	31.8	5.7	4715	8	US-08-781-986A-203	Sequence 203, App
18	31.8	5.7	4715	13	US-10-328-624-203	Sequence 203, App
19	31.6	5.7	521	13	US-10-424-509-50022	Sequence 50022, A
20	31.6	5.7	811	15	US-10-184-644-414	Sequence 414, App
21	31.6	5.7	811	15	US-10-184-644-414	Sequence 414, App
22	31.6	5.7	17286	9	US-09-764-877-3234	Sequence 3234, App
23	31.6	5.7	17286	16	US-10-243-515-3234	Sequence 3234, App
24	31.2	5.6	612	13	US-10-027-632-125951	Sequence 125951, App
25	31.2	5.6	612	13	US-10-027-632-125952	Sequence 125951, App
26	31.2	5.6	612	16	US-10-027-632-125951	Sequence 125951, App
27	31.2	5.6	612	16	US-10-027-632-125952	Sequence 125952, App
28	31.2	5.6	1300	10	US-09-934-455-251	Sequence 251, App
29	31.2	5.6	1300	13	US-10-225-066A-841	Sequence 841, App
30	31.2	5.6	1300	16	US-10-374-780A-309	Sequence 309, App
31	31.2	5.6	1308	13	US-10-282-122A-14107	Sequence 14107, A
32	31.2	5.6	11485	15	US-10-205-219-144	Sequence 144, App
33	31.2	5.6	15275	9	US-09-764-869-1475	Sequence 1475, App
34	31.2	5.6	15275	15	US-10-091-504-1475	Sequence 1475, App
35	31.2	5.6	15275	16	US-10-027-577-1475	Sequence 1475, App
36	31.2	5.6	856	13	US-10-027-632-5858	Sequence 5858, App
37	31.2	5.6	856	16	US-10-027-632-5858	Sequence 5858, App
38	31.2	5.6	918	9	US-09-452-599-162	Sequence 162, App
39	31.2	5.6	918	15	US-10-121-120-162	Sequence 162, App
40	30.8	5.5	105184	10	US-09-847-513A-1	Sequence 1, Appl1
41	30.8	5.5	33248	13	US-10-672-787-24	Sequence 24, Appl1
42	30.6	5.5	476	9	US-09-864-761-5512	Sequence 5512, App
43	30.6	5.5	659	13	US-10-027-632-152659	Sequence 152659, App
44	30.6	5.5	659	16	US-10-027-632-152659	Sequence 152659, App
45	30.6	5.5	1437	13	US-10-282-122A-32888	Sequence 32888, A

## ALIGNMENTS

RESULT 1

US-10-603-260-4

Sequence 4, Application US/10603260

Publication No. US20040009570A1

GENERAL INFORMATION:

APPLICANT: CJ Corporation

TITLE OF INVENTION: An alkaline lipase from *Vibrio metchnikovii* RH530 and a

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/10/603,260

CURRENT FILING DATE: 2003-06-24

NUMBER OF SEQ ID NOS: 5

SOFTWARE: Koparentin 1.71

SEQ ID NO 4

LENGTH: 555

TYPE: DNA

ORGANISM: *Vibrio metchnikovii* RH530

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(555)

OTHER INFORMATION: val12 gene

US-10-603-260-4

Query Match 100.0%; Score 555; DB 16; Length 555;  
Best Local Similarity 100.0%; Pred. No. 1.6e-179;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGATTATCTGTTGATGACTCTATGATGCTGTTGATGATGCTGCTAT 60

DB 1 ATGCAGATTATCTGTTGATGACTCTATGATGCTGTTGATGATGCTGCTAT 60

QY CATCGCTGATTAATGAGTATGCTACTCAACATAGCTAGCACTGCTATC 120

DB CATCGCTGATTAATGAGTATGCTACTCAACATAGCTAGCACTGCTATC 120

QY 121 GATGATGAGGATTTTGGCGCTTGAACGATGCTGCTATGCTGCTGCTATGCT 180

Db	121	GATATGAGGGCAATTTTTCGGCGCCTTGACCGATGGCTCACTCAATGCTCGCTTAAGCT	180
OY	181	TTAGTCGGAACAAGTTTGGGCGGAGTTGGTGAATCAAAAGTTATCTAAGATCGCGGCAACG	240
Db	181	TTAGTCGGAACAAGTTTGGGCGGAGTTGGTGAATCAAAAGTTATCTAAGATCGCGGCAACG	240
OY	241	TCTGTGAAACCCCTCTCCCATGTGTCGCGCATTCGGCTCACTTTGGCAAGAGCTTCCATT	300
Db	241	TCTGTGAAACCCCTCTCCCATGTGTCGCGCATTCGGCTCACTTTGGCAAGAGCTTCCATT	300
OY	301	GTCATATAAATTGAGCAATTAGGTTTAAAGGCTGCGACTAAGTAAATTCAGCAAAATTTGGG	360
Db	301	GTCATATAAATTGAGCAATTAGGTTTAAAGGCTGCGACTAAGTAAATTCAGCAAAATTTGGG	360
OY	361	TTAAAGAACAACGACGACGAATCCCGCATTCACAAAAATCAGGACGATTTGACAGAAACG	420
Db	361	TTAAAGAACAACGACGACGAATCCCGCATTCACAAAAATCAGGACGATTTGACAGAAACG	420
OY	421	ATACCTTTAGGAGCTGCGCAGCCTTTTACTGCGCGATTCACCTGACTCCGATGATACGATC	480
Db	421	ATACCTTTAGGAGCTGCGCAGCCTTTTACTGCGCGATTCACCTGACTCCGATGATACGATC	480
OY	481	ACAGTAGAAGAAACCAAAATAGTGGATGACAGATCATATCGAGATTCACACCACTTCA	540
Db	481	ACAGTAGAAGAAACCAAAATAGTGGATGACAGATCATATCGAGATTCACACCACTTCA	540
OY	541	TACGGAATGCTGTT	555
Db	541	TACGGAATGCTGTT	555

RESULT 2  
 US-10-603-260-1  
 ; Sequence 1, Application US/10603260  
 ; Publication No. US20040009570A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CJ Corporation  
 ; TITLE OF INVENTION: An alkaline lipase from *Vibrio metchnikovii* RH530 and a  
 ; TITLE OF INVENTION: nucleotide sequence encoding the same  
 ; FILE REFERENCE:  
 ; CURRENT APPLICATION NUMBER: US/10/603,260  
 ; CURRENT FILING DATE: 2003-06-24  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: Koparentin 1.71  
 ; SEQ ID NO 1  
 ; LENGTH: 2578  
 ; TYPE: DNA  
 ; ORGANISM: *Vibrio metchnikovii* RH530  
 US-10-603-260-1

	Query Match	Similarity	100.0%	Score 555	DB 16	length 2578				
	Best Local	Similarity	100.0%	Pred. No. 4e-179						
	Matches	555	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	1	ATGCAGATTATCTTGTGCTATGAGCTCTATATGCAATGCGCTGTGTGATGATGATCCGCTTAGT	60							
DB	1561	ATGCAGATTATCTTGTGCTATGAGCTCTATATGCAATGCGCTGTGTGATGATGATCCGCTTAGT	1620							
QY	61	CATGATCGCAATAATGGGTATGCTACTCAACCATTAGTCAACATCACTACGCTATC	120							
DB	1631	CATGATCGCAATAATGGGTATGCTACTCAACCATTAGTCAACATCACTACGCTATC	1680							
QY	121	GATGATGAGCCATTTTTCGCCGCTTGACGATGCTGCTCATAGCTCGGCTTAAGCT	180							
DB	1681	GATGATGAGCCATTTTTCGCCGCTTGACGATGCTGCTCATAGCTCGGCTTAAGCT	1740							
QY	181	TTAGTCGACACAGTTTGCGCGATGTGTGATCAACGTTATCTGAATTCGCGCACCG	240							
DB	1741	TTAGTCGACACAGTTTGCGCGATGTGTGATCAACGTTATCTGAATTCGCGCACCG	1800							
QY	241	TCCTGTGAACCTCTCCCATGTCGTGCGCAGTCGCTCACTTTTGCAAGAGAGCTTCATT	300							
DB	1801	TCCTGTGAACCTCTCCCATGTCGTGCGCAGTCGCTCACTTTTGCAAGAGAGCTTCATT	1860							

QY	361	GTCAATAAAAATTGAGCAATTAGGTTTATGGGGTGGCACTAGGTAATTTCAGACGAATTTGGG	360
Db	1861	GTCAATAAAAATTGAGCANTTAGGTTTATGGGGTGGCACTAGGTAATTTCAGACGAATTTGGG	1920
QY	361	TTAAAGAACACGACGACGAATCCCGCTATCCACAAAATACGGACGATATTGCAGAACG	420
Db	1921	TTAAAGAACACGACGACGAATCCCGCTATCCACAAAATACGGACGATATTGCAGAACG	1980
QY	421	ATTACCTTTAGGGCTGCGCAGGCTTTTATCTGGCGCATCCACTGAGCTCGATGTAACGATC	480
Db	1991	ATTACCTTTAGGGCTGCGCAGGCTTTTATCTGGCGCATCCACTGAGCTCGATGTAACGATC	2040
QY	481	ACAGTAGAAGAAACCAAAATAGCTGGCATGACAGATCATATCGGATATCCACCACTTCA	540
Db	2041	ACAGTAGAAGAAACCAAAATAGCTGGCATGACAGATCATATCGGATATCCACCACTTCA	2100
QY	541	TACGAGAAATGCTGTT	555
Db	2101	TACGAGAAATGCTGTT	2115

```

RESULT 3
US-10-221-714A-473/c
; Sequence 473. Application US/10221714A
; Publication No. US20040048254A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with
FILE REFERENCE: 5013.1005
CURRENT APPLICATION NUMBER: US/10/221.714A
CURRENT FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: PCT/EP01/02955
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: DE 10013847.0
PRIOR FILING DATE: 2000-03-15
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 540
SEQ ID NO 473
; LENGTH: 26997
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-473

```

Query Match	6.6%	Score 36.8	DB 13	Length 26997
Best Local Similarity	47.1%	Pred. No. 1.47	127	Indels 0
Matches 113	Conservative 0	Mismatches 0	Gaps 0	
QY	277	TCACCTTGCAGAGAGCTTCATATGTCATATTAATTAATGAGCAATTAAGTTAGGGGTGGCA	336	
Db	13416	TCTCATGAGAACTTACACATCCATCATATTAATTAATTAATTAATTAATTAATTAATTAATTC	1335	
QY	337	CTAGGTAAATTCAGAGAAATTTGGGTTAAAGAACAGACAGCAATCCGCTATCCACAA	396	
Db	13356	CTACACATTTATTAACCTTGCACCTATATCCACTGGAAACCAATACCTCTTTAAACAA	1329	
QY	397	AAATCAGGACGATATTGCAGAACATACCTTTAGGGCTGGCAGCCCTTTTACTGGCGAT	456	
Db	13296	ATTAATAATTAATACTACAAAAACGCTACCTTATATCTACAAATAATATCCCGACCTCTAC	1323	
QY	457	CCACTGGACCTCCGATGGTACCGTACAGTGAAGAAACCAAAATTAAGCTGCATGACAGAT	516	

Db 13236 CCACCAACCAACCAATTAACCTCCCACTATTAACCAACCAACCAACCAATTAACCAAT 13177

RESULT 4

US-10-087-192-1135

Sequence 1135, Application US/10087192

Publication No. US20020182586A1

GENERAL INFORMATION:

APPLICANT: Morris, David W.

APPLICANT: Engelhard, Eric K.

TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: CANCER

FILE REFERENCE: 529452000122

CURRENT APPLICATION NUMBER: US/10/087,192

CURRENT FILING DATE: 2002-03-01

PRIOR APPLICATION NUMBER: US 09/747,377

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: US 09/798,586

PRIOR FILING DATE: 2001-03-02

NUMBER OF SEQ ID NOS: 2059

SOFTWARE: fastseq for windows Version 4.0

SEQ ID NO 1135

LENGTH: 32185

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)...(32185)

OTHER INFORMATION: n = A,T,C or G

US-10-087-192-1135

Query Match

Best Local Similarity 6.0%; Score 33.2; DB 13; Length 32185;

Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 380 AATCCCGCTATCCACAAATATGAGGAGATTTGACAGAGATACCTTTAGGCTGCGCA 439

Db 31738 AATCAACCCCTTCTCCCAAGATTTGACATTTGAAATCACTCGTCTCGTAAGAT 31797

QY 440 GCGTTTACTGCGCATCTGCACTCGCATCGATGTCACCTGACATGAGAAACCAAA 499

Db 31798 GCGTTTACTGCGCATCTGCACTCGCATCGATGTCACCTGACATGAGAAACCAAA 499

QY 500 TAGCTGCAATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 553

Db 31858 GTCATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 31911

RESULT 5

US-09-783-590-1460

Sequence 1460, Application US/09783590

Patent No. US20020110850A1

GENERAL INFORMATION:

APPLICANT: Dillon, Patrick J.

APPLICANT: Haseltine, William A.

APPLICANT: Li, Haodong

APPLICANT: Rosen, Craig A.

APPLICANT: Ruben, Steven M.

TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2

FILE REFERENCE: PO-16.2C1

CURRENT APPLICATION NUMBER: US/09/783,590

CURRENT FILING DATE: 2000-02-15

PRIOR APPLICATION NUMBER: 08/420,856

PRIOR FILING DATE: 1995-04-12

PRIOR APPLICATION NUMBER: 08/346,731

PRIOR FILING DATE: 1994-11-21

NUMBER OF SEQ ID NOS: 12485

SOFTWARE: Patencin Ver. 2.0

SEQ ID NO 1460

LENGTH: 435

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (142)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (146)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (148)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (166)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (333)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (335)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (374)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (379)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (388)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (401)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (407)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (410)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc\_feature

LOCATION: (414)

OTHER INFORMATION: n equals a,t,g, or c

US-09-783-590-1460

Query Match

Best Local Similarity 5.9%; Score 33; DB 9; Length 435;

Matches 99; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 47 TGATCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 106

Db 117 TGATCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 176

QY 107 ACTCACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 166

Db 177 GGCCTAAACAGCTCTTAAATATATTTTATGAGGAGGCTTCAATGGAATGATGCTG 236

QY 167 CCTCGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 226

Db 237 TAAATCCGCACTTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 296

QY 227 AATCGCGGCAACCGCTCTGTAACCCCTCTCC 259

Db 297 ACCAGCTGCGCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 329

RESULT 6

US-10-087-192-2011/c

Sequence 2011, Application US/10087192

Publication No. US20020182586A1

GENERAL INFORMATION:

APPLICANT: Morris, David W.

APPLICANT: Engelhard, Eric K.

TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: CANCER

FILE REFERENCE: 529452000122



```

1 CURRENT FILING DATE: 2003-02-20
2 PRIOR APPLICATION NUMBER: 60/191,078
3 PRIOR FILING DATE: 2000-03-31
4 PRIOR APPLICATION NUMBER: 60/206,848
5 PRIOR FILING DATE: 2000-05-23
6 PRIOR APPLICATION NUMBER: 60/220,727
7 PRIOR FILING DATE: 2000-05-26
8 PRIOR APPLICATION NUMBER: 60/230,335
9 PRIOR FILING DATE: 2000-09-06
10 PRIOR APPLICATION NUMBER: 60/230,347
11 PRIOR FILING DATE: 2000-09-09
12 PRIOR APPLICATION NUMBER: 60/242,578
13 PRIOR FILING DATE: 2000-10-23
14 PRIOR APPLICATION NUMBER: 60/253,625
15 PRIOR FILING DATE: 2000-11-27
16 PRIOR APPLICATION NUMBER: 60/257,931
17 PRIOR FILING DATE: 2000-12-22
18 PRIOR APPLICATION NUMBER: 60/267,636

```

Qy	274	GGCTCACTTTGGAAGAGCTTCCATTTGTCAATTAATATGACAAATAGCTTTAGGGCTG	333
Db	1602662	GGCTCACTTTCARATTTCTTCTCTGATCCCTTGAAATCAGCATTTTTTTTTTCAAGAG	1602603
Qy	334	GCACTAGTAATTCAGCGAAATTTGGGTAAAGAACACGACGACGAATCCCGCTATCCA	393
Db	1602602	TCTTGTTCTCTTTTACCGAAGAAATGGATTTTGAAGAACAAATATGAGGGCTTAAGTGTCT	1602544
Qy	394	CAAAATCAGGACAGTATTCAGGAAACATACCTTTAGGGCTG	435
Db	1602542	CATTGCTACTGGGTATGTAGCTTTCAGACCTCTCGGGCTG	1602501

## RESULT 9

```
US-10-419-723-1/c
; Sequence 1, Application US/10419723
; Publication No. US20040014099a1
; GENERAL INFORMATION:
; APPLICANT: Gretaodotitir, Solveig
; APPLICANT: Jonadotitir, Sif
; APPLICANT: Reynadotitir, Sigridur Th.
; APPLICANT: Thorleifsson, Gudmar
; APPLICANT: Gulcher, Jeffrey
; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
; FILE REFERENCE: 2345.2010-005
; CURRENT APPLICATION NUMBER: US/10/419,723
; PRIOR FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 10/255,120
; PRIOR FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 10/067,514
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 09/811,352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-419-723-1

Query Match      5.8%; Score 32.4; DB 16; Length 1691139;
Best Local Similarity 50.0%; Pred. No. 4.9e+02;
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 274 GGCCTACCTTGCAAGAGAGCTTCCTCATTTGCAATTAATAGCATTAGGTGAGGGGTG 333
Db 1602662 GGCCTACCTTGCAATTAATTTCTTCTGAGTCCCTGAATATCGATTTTTCCTAAAGAG 1602603

Qy 334 GCACCTAGTAATTCAGCAATTTGGGTTAAAGAACGACGACGATCCCGCTATCCA 393
Db 1602602 TCTTGTCCTTTTACCAAGAAATGGTATTAAGAAACCAATGAGGTGCTTAAGTGTGCT 1602543

Qy 394 CAAATAGGAGGAGTATTCAGCAAGCATATCTTAGGGGTG 435
Db 1602542 CATGCTACTGGGGTATGATGCTTTTAGACCTCTCGGCTG 1602501

RESULT 10
US-10-087-192-226/c
; Sequence 226, Application US/10087192
; Publication No. US20020182586a1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226
; LENGTH: 653122
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(653122)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-226
```

```
Query Match      5.8%; Score 32.2; DB 13; Length 653122;
Best Local Similarity 61.2%; Pred. No. 3.5e+02;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 362 TAAAGACACGACGACGATCCGCTATCCAAATAATCAGGAGTATTGACGACGACGA 421
Db 142345 TTAAGTACACGATAGCAATCTTCGATAGGCTAAACATCAGGACGAGACGAGTAAAA 142286

Qy 422 TACCTTAGGCTGGCGACCTTTT 446
Db 142285 TAGCTTAGGATACCACTATTTT 142261

RESULT 11
US-10-282-122A-18072
; Sequence 18072, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,076
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PAM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 18072
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
US-10-282-122A-18072

Query Match      5.8%; Score 32; DB 13; Length 549;
Best Local Similarity 65.3%; Pred. No. 6.3; 25; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 332 TGGCACTAGTAATTCAGCAATTTGGGTTAAAGAACGACGACGATCCCGCTATC 391
Db 428 TGAATCTTGGAATCTCTGAAATTTTGTCCGATTAACGATGAAGAGTCAGTTTC 487

Qy 392 CACAAATATCAG 403
```

Db 488 TAAACAATCCG 499

## RESULT 12

US-09-938-842A-1790  
Sequence 1790, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SCRIPI300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 1790  
LENGTH: 969  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1790

Query Match  
Best Local Similarity 54.8%; Pred. No. 10;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 89 CTCMAACCTTAGCTACACTCACTCGCTATCATGATGAGGCCATTTTGGCGGCTTG 148  
Db 392 CTCMAACCAACCAACCAACCTCCCTGTTCTCCCTGGAAGACGGTGTCCAGCACAA 451  
Qy 149 ACCGATCGCTCACTCACTGCTGCTTATGCTTTAGTGGACACAGTTGGCGG 203  
Db 452 AACGCCCTGCTCCACTGTCAGATTAAGATGAAGTCAAGACATGTGTGCGG 506

## RESULT 13

US-09-934-455-271  
Sequence 271, Application US/09934455  
GENERAL INFORMATION:  
APPLICANT: Adam, Luc  
APPLICANT: Creelman, Robert  
APPLICANT: Dubell, Arnold  
APPLICANT: Heard, Jacqueline  
APPLICANT: Jiang, Cai-Zhong  
APPLICANT: Keddie, James  
APPLICANT: Pilgrim, Marsha  
APPLICANT: Ratcliffe, Oliver  
APPLICANT: Reuber, Lynne  
APPLICANT: Riechmann, Jose Luis  
APPLICANT: Yu, Guo-Liang  
APPLICANT: Pineta, Omaira  
TITLE OF INVENTION: Genes for Modifying Plant Traits IV  
FILE REFERENCE: MBI-0025  
CURRENT APPLICATION NUMBER: US/09/934,455  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/227439  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: MBI-0022  
PRIOR FILING DATE: 2001-11-16  
PRIOR APPLICATION NUMBER: MBI-0023  
PRIOR FILING DATE: 2001-04-17  
NUMBER OF SEQ ID NOS: 516  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 271

LENGTH: 969  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(969)  
OTHER INFORMATION: G1835  
US-09-934-455-271

Query Match  
Best Local Similarity 54.8%; Pred. No. 10;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 89 CTCMAACCTTAGCTACACTCACTCGCTATCATGATGAGGCCATTTTGGCGGCTTG 148  
Db 392 CTCMAACCAACCAACCAACCTCCCTGTTCTCCCTGGAAGACGGTGTCCAGCACAA 451  
Qy 149 ACCGATCGCTCACTCACTGCTGCTTATGCTTTAGTGGACACAGTTGGCGG 203  
Db 452 AACGCCCTGCTCCACTGTCAGATTAAGATGAAGTCAAGACATGTGTGCGG 506

## RESULT 14

US-09-938-842A-1790  
Sequence 1790, Application US/09938842A  
Publication No. US20040009476A9  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SCRIPI300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 1790  
LENGTH: 969  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1790

Query Match  
Best Local Similarity 54.8%; Pred. No. 10;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 89 CTCMAACCTTAGCTACACTCACTCGCTATCATGATGAGGCCATTTTGGCGGCTTG 148  
Db 392 CTCMAACCAACCAACCAACCTCCCTGTTCTCCCTGGAAGACGGTGTCCAGCACAA 451  
Qy 149 ACCGATCGCTCACTCACTGCTGCTTATGCTTTAGTGGACACAGTTGGCGG 203  
Db 452 AACGCCCTGCTCCACTGTCAGATTAAGATGAAGTCAAGACATGTGTGCGG 506

## RESULT 15

US-10-225-066A-377  
Sequence 377, Application US/10225066A  
Publication No. US20030226173A1  
GENERAL INFORMATION:  
APPLICANT: Mendel Biotechnology, Inc.  
APPLICANT: Ratcliffe, Oliver  
APPLICANT: RIECHMANN, Jose Luis  
APPLICANT: ADAM, Luc J  
APPLICANT: DUBELL, Arnold T  
APPLICANT: HEARD, Jacqueline E

```

; APPLICANT: PILGRIM, Marsha L
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: REUBER, T. Lynne
; APPLICANT: CREELMAN, Robert A
; APPLICANT: PINEDA, Omaira
; APPLICANT: YU, Guo-Liang
; APPLICANT: BROUN, Pierre E
; TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MB10036-2 US
; CURRENT APPLICATION NUMBER: US/10/225,066A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 09/837,444
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 1122
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 377
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-225-066A-377

```

```

Query Match      5.7%; Score 31.8; DB 13; Length 969;
Best Local Similarity 54.8%; Pred. No. 10;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY      89  CTCAACCATTAAGCTACACTCACTCGCTATCGATGATGAGGCCATTTCGGCCCTTG 148
DB      392 CTCAACCAACCAACCAACCTCCCTTGTCTCCCTGGAAGCAGGTCGTCACGACAA 451

QY      149  ACCGATGCTCTACTCATGCTCGCTTAATGCTTTAGTCGACACAGTTGGGCGG 203
DB      452 AACGCCCTGCTCCACGCTGTCAGGATAAAGATAGATCAAGACAATGTGTGCGG 506

```

Search completed: April 30, 2004, 23:19:59  
 Job time : 261.561 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 11:16:06 ; Search time 57 Seconds  
(without alignments)  
917.040 Million cell updates/sec

Title: US-10-603-260-5  
Perfect score: 936  
Sequence: 1 MQLIVHGLVHGLVHPLS.....KIAMGDHIAISTSYENAV 185

Scoring table: BLASTN62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseqp1980s:\*\n2: Geneseqp1990s:\*\n3: Geneseqp2000s:\*\n4: Geneseqp2001s:\*\n5: Geneseqp2002s:\*\n6: Geneseqp2003as:\*\n7: Geneseqp2003bs:\*\n8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323.5	34.6	201	2	AAW23078 LA11.1 es
2	98	10.5	323	3	AAW20858 Arabidops
3	98	10.5	339	3	AAW20857 Arabidops
4	98	10.5	340	3	AAW20856 Arabidops
5	97	10.4	277	1	AAW20065 Sequence
6	97	10.4	277	2	AAW48085 P. fragi
7	96.5	10.3	453	6	ADA33782 Actinocba
8	96	10.3	316	7	ADG94967 E. faeciu
9	96	10.3	319	1	AAW91951 Polypepti
10	96	10.3	319	2	AAW88013 Mature Ps
11	96	10.3	319	2	AAW88015 Mature Ps
12	96	10.3	319	2	AAW88018 Mature Ps
13	96	10.3	319	2	AAW88009 Mature Ps
14	96	10.3	319	2	AAW88014 Mature Ps
15	96	10.3	319	2	AAW88016 Mature Ps
16	96	10.3	319	2	AAW88017 Mature Ps
17	96	10.3	319	2	AAW09624 Pseudomon
18	96	10.3	319	2	AAW09625 Pseudomon
19	96	10.3	319	7	ADD13298 Pseudomon
20	96	10.3	339	2	AAW52656 E. coli Om
21	96	10.3	358	2	AAW10614 Pseudomon
22	96	10.3	358	2	AAW20177 P. glumae
23	96	10.3	360	2	AAW10438 Pseudomon
24	96	10.3	361	2	AAW10437 Pseudomon
25	96	10.3	631	4	ABG27846 Novel hum

26	94	10.0	319	2	AAW88010 Mature Ps
27	93.5	10.0	319	2	AAW88012 Mature Ps
28	93.5	10.0	319	2	AAW88011 Mature Ps
29	93.5	10.0	364	2	AAW77245 P. cepaci
30	92	9.8	289	2	AAW53933 Pseudomon
31	91.5	9.8	262	2	AAW23073 Thermococ
32	91.5	9.8	262	6	ABU56926 Thermococ
33	91.5	9.8	262	7	ABG75902 Thermococ
34	91.5	9.8	262	7	AAW38797 Thermococ
35	91	9.7	313	2	AAW85077 Pseudomon
36	91	9.7	320	2	AAW87918 Pseudomon
37	91	9.7	320	2	AAW09620 Pseudomon
38	91	9.7	320	3	AAW45743 Arabidops
39	91	9.7	336	3	AAW45742 Arabidops
40	91	9.7	337	3	AAW45741 Arabidops
41	91	9.7	364	2	AAW34633 Lipase B
42	90.5	9.7	320	2	AAW09622 Pseudomon
43	90.5	9.7	363	2	AAW61276 Burkholde
44	90.5	9.7	363	2	AAW6101 E. coli e
45	90.5	9.7	363	4	AAW73542 Burkholde

## ALIGNMENTS

RESULT 1

ID AAW23078 standard; protein; 201 AA.

AC AAW23078;

DT 27-AUG-2003 (revised)

DT 17-FEB-1998 (first entry)

XX LA11.1 esterase es2.

XX Esterase; thermostable enzyme; ester; chiral compound; cheese; pulp;

XX paper; lignin removal; sugar; lignocellulose; disease resistance;

XX feedstuff.

XX Unidentified.

XX WO9730160-A1.

XX 21-AUG-1997.

XX 11-FEB-1997; 97WO-US002039.

XX 16-FEB-1996; 96US-00602359.

XX (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX Robertson DE, Murphy D, Reid J, Maffia AM, Link S, Swanson RV;

XX Warren PV, Kosmicka A, Callen W,

XX WPI, 1997-425035/39.

XX N-PSDB; AAT79331.

XX Nucleic acid encoding heat stable esterase from thermophilic bacteria -

XX which is active in organic solvents, useful in cheese or paper

XX manufacture, and to study plant resistance to disease.

XX Disclosure; Page 63-64; 113pp; English.

XX This protein comprises LA11.1 esterase es2. Newly identified

XX polynucleotides (AAT79331-40) encoding esterases (AAW23069-88), some of

XX which are claimed, can be used for recombinant production of the enzymes

XX in host cells. The esterases are stable at high temperature and in

XX organic solvents, making them superior for use in production of pure

XX chiral compounds used in pharmaceutical, agricultural and other chemical

XX industries. A method is claimed for transferring an amino group from an

XX amino acid to an alpha-keto acid using a claimed esterase. The enzymes

XX may also be useful as ripening starters in cheese making, in lignin



PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 04-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148311P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160747P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161932P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 10.5%; Score 98; DB 3; Length 323;  
Best Local Similarity 24.0%; Pred. No. 0.05;  
Matches 41; Conservative 18; Mismatches 44; Indels 68; Gaps 7;

QY 3 IILVHGLY-----MHGLVMEPLSHRLHKGYRQTTSYNSLAIDDEAFIRLRSLTH 55  
Db 55 LVFVHSTHAACWAEHLPPRSS-----GFDSYANSLDQGSDEPLGTVAGTQTH 108  
QY 56 A-----SPNALVGHSLGLVYIKRYLESRAPECTLSHVVAIGSPLOQASIVN 102  
Db 109 ASDIADFIRESNLGSSPPVLIGHSPGLIVQYLL-----ANIVN 146  
QY 103 KIEQLGLGVALGNSAEPLGKHHDESRYPQKGS--IAGTIPLGRLSLLR 151  
Db 147 K-QSLG-----TENAPPELSGAVWCVSPPSGNSGLVLR 179

## RESULT 3

ID AAG20857 standard; protein; 339 AA.

XX AAG20857;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 23202.

XX Protein identification; signal transduction pathway; metabolic pathway;

XX hybridization assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

XX 09-MAR-1999; 99US-0123548P.

XX 23-MAR-1999; 99US-0125788P.

XX 25-MAR-1999; 99US-0126264P.

XX 29-MAR-1999; 99US-0126785P.

XX 01-APR-1999; 99US-0127462P.

XX 06-APR-1999; 99US-0128234P.

XX 08-APR-1999; 99US-0128714P.

XX 16-APR-1999; 99US-0129845P.

XX 19-APR-1999; 99US-0130077P.

XX 21-APR-1999; 99US-0130449P.

XX 23-APR-1999; 99US-0130510P.

XX 28-APR-1999; 99US-0130891P.

XX 30-APR-1999; 99US-0133404P.

XX 30-APR-1999; 99US-0133407P.

XX 04-MAY-1999; 99US-0133484P.

XX 05-MAY-1999; 99US-0133485P.

XX 06-MAY-1999; 99US-0133486P.

XX 07-MAY-1999; 99US-0133487P.

XX 11-MAY-1999; 99US-0133863P.

XX 14-MAY-1999; 99US-0134256P.

XX 14-MAY-1999; 99US-0134218P.



PR	14-MAY-1999	99US-013442199
PR	14-MAY-1999	99US-01344221P
PR	14-MAY-1999	99US-01344230P
PR	18-MAY-1999	99US-0134768R
PR	18-MAY-1999	99US-01349411P
PR	20-MAY-1999	99US-0135124P
PR	21-MAY-1999	99US-0135533P
PR	24-MAY-1999	99US-0136522P
PR	27-MAY-1999	99US-0136392P
PR	28-MAY-1999	99US-0136782P
PR	01-JUN-1999	99US-0137222P
PR	03-JUN-1999	99US-0137528P
PR	04-JUN-1999	99US-0137502P
PR	07-JUN-1999	99US-0137724P
PR	08-JUN-1999	99US-0138094P
PR	10-JUN-1999	99US-0138540P
PR	10-JUN-1999	99US-0138847P
PR	14-JUN-1999	99US-0139119P
PR	15-JUN-1999	99US-0139452P
PR	16-JUN-1999	99US-0139453P
PR	17-JUN-1999	99US-0139492P
PR	18-JUN-1999	99US-0139454P
PR	18-JUN-1999	99US-0139455P
PR	18-JUN-1999	99US-0139456P
PR	18-JUN-1999	99US-0139457P
PR	18-JUN-1999	99US-0139458P
PR	18-JUN-1999	99US-0139459P
PR	18-JUN-1999	99US-0139460P
PR	18-JUN-1999	99US-0139461P
PR	18-JUN-1999	99US-0139462P
PR	23-JUN-1999	99US-0139463P
PR	23-JUN-1999	99US-0140053P
PR	24-JUN-1999	99US-0140695P
PR	24-JUN-1999	99US-0140823P
PR	28-JUN-1999	99US-0140931P
PR	30-JUN-1999	99US-0141287P
PR	01-JUL-1999	99US-0141642P
PR	01-JUL-1999	99US-0142154P
PR	02-JUL-1999	99US-0142055P
PR	06-JUL-1999	99US-0142390P
PR	08-JUL-1999	99US-0142803P
PR	09-JUL-1999	99US-0142920P
PR	12-JUL-1999	99US-0142972P
PR	13-JUL-1999	99US-0143542P
PR	14-JUL-1999	99US-0143624P
PR	15-JUL-1999	99US-0144005P
PR	16-JUL-1999	99US-0144086P
PR	16-JUL-1999	99US-0144088P
PR	19-JUL-1999	99US-0144335P
PR	19-JUL-1999	99US-0144332P
PR	20-JUL-1999	99US-0144632P
PR	20-JUL-1999	99US-0144884P
PR	21-JUL-1999	99US-0144814P
PR	21-JUL-1999	99US-0145086P
PR	21-JUL-1999	99US-0145088P
PR	22-JUL-1999	99US-0145085P
PR	22-JUL-1999	99US-0145087P
PR	22-JUL-1999	99US-0145089P
PR	22-JUL-1999	99US-0145192P
PR	22-JUL-1999	99US-0145145P
PR	23-JUL-1999	99US-0145521P
PR	23-JUL-1999	99US-0145524P

PR	27-JUN-1999	99US-0145376P
PR	27-JUN-1999	99US-0145313P
PR	27-JUN-1999	99US-0145318P
PR	28-JUN-1999	99US-0145519P
PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146386P
PR	02-AUG-1999	99US-0146389P
PR	02-AUG-1999	99US-0146388P
PR	04-AUG-1999	99US-0147704P
PR	04-AUG-1999	99US-0147702P
PR	05-AUG-1999	99US-0147392P
PR	10-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-0148341P
PR	12-AUG-1999	99US-0148341P
PR	13-AUG-1999	99US-0148656P
PR	16-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149175P
PR	18-AUG-1999	99US-0149462P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149723P
PR	20-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149902P
PR	25-AUG-1999	99US-0149930P
PR	25-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151066P
PR	27-AUG-1999	99US-0151080P
PR	30-AUG-1999	99US-0151303P
PR	31-AUG-1999	99US-0151348P
PR	01-SEP-1999	99US-0151530P
PR	07-SEP-1999	99US-0153263P
PR	10-SEP-1999	99US-0153070P
PR	13-SEP-1999	99US-0153758P
PR	15-SEP-1999	99US-0154018P
PR	16-SEP-1999	99US-0154039P
PR	20-SEP-1999	99US-0156458P
PR	20-SEP-1999	99US-0156569P
PR	04-OCT-1999	99US-0157117P
PR	05-OCT-1999	99US-0157753P
PR	06-OCT-1999	99US-0157865P
PR	07-OCT-1999	99US-0158032P
PR	08-OCT-1999	99US-0158283P
PR	12-OCT-1999	99US-0158369P
PR	13-OCT-1999	99US-0159293P
PR	13-OCT-1999	99US-0159284P
PR	13-OCT-1999	99US-0159289P
PR	14-OCT-1999	99US-0159292P
PR	14-OCT-1999	99US-0159331P
PR	14-OCT-1999	99US-0159637P
PR	14-OCT-1999	99US-0159658P
PR	18-OCT-1999	99US-0159584P
PR	21-OCT-1999	99US-0160747P
PR	21-OCT-1999	99US-0160768P
PR	21-OCT-1999	99US-0160770P
PR	21-OCT-1999	99US-0160814P
PR	21-OCT-1999	99US-0160815P
PR	22-OCT-1999	99US-0160981P
PR	22-OCT-1999	99US-0160982P
PR	22-OCT-1999	99US-0160988P
PR	22-OCT-1999	99US-0160989P

PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161358P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161932P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 10.5%; Score 98; DB 3; Length 339;  
Best Local Similarity 24.0%; Pred. No. 0.053;  
Matches 41; Conservative 18; Mismatches 44; Indels 68; Gaps 7;

QY 3 ILVHGLY-----MGLVHNPISHRHLKGYTQIINYSNLALIDEAIFRRLDRLTLH 55  
DB 71 LVFVHGSYHAMCMNHEWLPFFSSS-----GFDYSVAVSLGGESEDEPIGTAGTQTH 124  
QY 56 A-----SPNALVGHSLGGLVYKRYLESRAPSCETLSHVVAIGSPLOGASIVN 102  
DB 125 ASDIADPIESNLGSPVTLIGHSPGLIVQYLL-----ANTVN 162  
QY 103 KIEQLGLGVALGNNAEFGKKEHDESKRYPKSGS--TAGTIPUGLRSLMLR 151  
DB 163 K-QSIG-----TENAFPELSGAVMVCVPPSGNSGLVLR 195

## RESULT 4

AA020856  
ID AA020856 standard; protein, 340 AA.

XX AA020856;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 23201.

XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126765P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130851P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135533P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139494P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139895P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 19-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.

PR 23-JUL-1999; 99US-0145224P.  
 PR 26-JUL-1999; 99US-0145276P.  
 PR 27-JUL-1999; 99US-0145913P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147192P.  
 PR 06-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147303P.  
 PR 09-AUG-1999; 99US-0147416P.  
 PR 09-AUG-1999; 99US-0147493P.  
 PR 09-AUG-1999; 99US-0147935P.  
 PR 10-AUG-1999; 99US-0148177P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149375P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149929P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 23-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 30-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 29-SEP-1999; 99US-0156458P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.

PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 28-OCT-1999; 99US-0162142P.

Query Match 10.5%; Score 98; DB 3; Length 340;  
 Best Local Similarity 24.0%; Pred. No. 0.053;  
 Matches 41; Conservative 18; Mismatches 44; Indels 68; Gaps 7;

QY 3 IIVHGLY-----MHGLVHPLSHRLKGYRQTSYNSLAIDDEAFRRLDRSLTH 55  
 DB 72 LVFVHGSTHAWCMKMHLPFRSS-----CFDSYAVSLGQGSDEPLGIVAGTLOTH 125  
 QY 56 A-----SPNALVGHSLGLVYKRYLESRAPSCEITLSHVVAIGSPLQASIVN 102  
 DB 126 ASDIADPIESNLGSPPLVIGHSFGGLIVQYLL-----ANIYV 163  
 QY 103 KTEQLGLVALGNSAEFLKHHDDSRYPKSGS--IAGTIPLGRLSLR 151  
 DB 164 K-QSLG-----TENAFELSGAVWCVSPPSGSLVLR 196

## RESULT 5

AAAP90065  
 ID AAP90065 standard; protein; 277 AA.

AC AAP90065;  
 XX 25-MAR-2003 (revised)  
 DT 01-NOV-1989 (first entry)  
 XX Sequence encoding a Pseudomonas lipase.  
 DE  
 XX  
 XX Lipase; Pseudomonas; Pseudomonas fragi.  
 KW  
 OS Pseudomonas fragi.  
 XX  
 PN EP318775-A.  
 XX  
 XX  
 PD 07-JUN-1989.  
 XX  
 XX 18-NOV-1988; 88EP-00119211.  
 PF  
 XX 03-DEC-1987; 87JP-00306638.  
 PR 20-MAY-1988; 88JP-00123672.  
 PR 27-JUL-1988; 88JP-00187684.  
 XX  
 XX (CHCC) CHISSO CORP.  
 PA  
 XX Aoyama S, Inouye S, Yoshida N;  
 PI WPI, 1989-166833/23.  
 DR  
 XX DNA coding for a lipase - obtd. by digestion of Pseudomonas genus  
 PT bacterium DNA and transformation of E. coli cells.  
 XX  
 XX Claim 1, Fig 4; 15pp; English.  
 PS  
 XX Amino acid sequence encoding a Pseudomonas lipase. Used as a lipid-  
 CC hydrolysing enzyme for oils and fats processing; in medicine, diagnosis,  
 CC in detergents and digestants. It can catalyse ester hydrolysis, ester  
 CC synthesis or ester conversion, in the prodn. of chemical prods. esp.  
 CC optically active cpds. The DNA can be obtd. from P. fragi and cloned and  
 CC expressed in E. coli. (Updated on 25-MAR-2003 to correct PR field.)  
 XX  
 SQ Sequence 277 AA;

Query Match 10.4%; Score 97; DB 1; Length 277;  
 Best Local Similarity 24.6%; Pred. No. 0.051;  
 Matches 51; Conservative 31; Mismatches 79; Indels 46; Gaps 11;

QY 3 ILLVGL-----YMGVLVHPLSHRLKLYRTQTISYNSLAIDDEA---IFRR 48  
 DB 11 ILLVGLFGFDRIGSHHYFHG-----IKQALNECG-ASVFPPIISAANDNEARGDQLKQ 64  
 QY 49 LD--RSLTHASPNALVGHSLGLVTKRYLESRAPECSCTLSHVVALGSPLOGASIVNKIE- 105  
 DB 65 IHNLRQVGAQRVNLIGHSQALT-ARYVAALAP--ELIASVTSVSGPNHSGELADRLRL 121  
 QY 106 -----QLGLGYALGNSAEFG--LKEHDESRYPQKSGSIAGTIPUGLSLLRDPDSDG 158  
 DB 122 AFVPRIGETVVAALTTSSFAFLSALSGHPLPQNA-----LNAL---NALTTDG 168  
 QY 159 TTVVETKLAGMTDHIATSTSYENAV 185  
 DB 169 VAAFNROYPOGLPDRWGGMGPQVNAV 195

## RESULT 6

AAR48085  
 ID AAR48085 standard; protein; 277 AA.

AC AAR48085;

DT 25-MAR-2003 (revised)  
 DT 19-JUL-1994 (first entry)

DE P. fragi lipase A.

XX Lipase A; thermostability; specific activity; mechanical pulp;  
 XX ester hydrolysis; ester synthesis; interesterification; acid pH; pitch;  
 XX paper-making; detergent; digestive enzyme; cystic fibrosis.

XX Pseudomonas fragi.

XX MO9401541-A1.

XX 20-JAN-1994.

XX 05-JUL-1993; 93WO-DK000225.

XX 06-JUL-1992; 92DK-00000888.

XX (NOVO) NOVO-NORDISK AS.

PI Svendsen A, Pathar S, Egel-Mitani M, Borch K, Clausen G;

PI Hansen M;

DR WPI; 1994-035050/04.

XX New lipase variants, partic. from *Candida antarctica* - have aminoacid  
 PT substitutions to increase specific activity against substrates compared  
 PT to parent lipase(s).

XX Claim 13; Page 41-42; 60pp; English.

XX This sequence represents the P. fragi lipase protein. This lipase has  
 CC high thermostability and is active at acid pH. A variant of the mature  
 CC lipase A in which Phe23 has been replaced by a Trp residue has increased  
 CC specific activity compared to the parent lipase. The lipase or variant  
 CC may be used in ester hydrolysis, ester synthesis or interesterification.  
 CC They can also be used for avoiding pitch trouble in the process for the  
 CC production of mechanical pulp or a paper-making process using mechanical  
 CC pulp. They can also be used in detergents or as a digestive enzyme, eg.  
 CC in the treatment of cystic fibrosis. (Updated on 25-MAR-2003 to correct  
 CC FN field.)

XX Sequence 277 AA;  
 SQ

Query Match 10.4%; Score 97; DB 2; Length 277;  
 Best Local Similarity 24.6%; Pred. No. 0.051;  
 Matches 51; Conservative 31; Mismatches 79; Indels 46; Gaps 11;

QY 3 ILLVGL-----YMGVLVHPLSHRLKLYRTQTISYNSLAIDDEA---IFRR 48  
 DB 11 ILLVGLFGFDRIGSHHYFHG-----IKQALNECG-ASVFPPIISAANDNEARGDQLKQ 64  
 QY 49 LD--RSLTHASPNALVGHSLGLVTKRYLESRAPECSCTLSHVVALGSPLOGASIVNKIE- 105  
 DB 65 IHNLRQVGAQRVNLIGHSQALT-ARYVAALAP--ELIASVTSVSGPNHSGELADRLRL 121  
 QY 106 -----QLGLGYALGNSAEFG--LKEHDESRYPQKSGSIAGTIPUGLSLLRDPDSDG 158  
 DB 122 AFVPRIGETVVAALTTSSFAFLSALSGHPLPQNA-----LNAL---NALTTDG 168  
 QY 159 TTVVETKLAGMTDHIATSTSYENAV 185  
 DB 169 VAAFNROYPOGLPDRWGGMGPQVNAV 195

## RESULT 7

ADA33782  
 ID ADA33782 standard; protein; 453 AA.

AC ADA33782;

DT 20-NOV-2003 (first entry)

DE Acinetobacter baumannii protein #943.

XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
 XX plant biocontrol agent.

XX Acinetobacter baumannii.

XX US6562958-B1.

XX 13-MAY-2003.

XX 04-JUN-1999; 99US-00328352.

XX 09-JUN-1998; 98US-0088701P.

XX (GENO-) GENOME THERAPEUTICS CORP.

PI Breton G, Bush D;

DR WPI; 2003-576092/54.

DR N-PSDB; ADA29656.

XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
 PT for diagnosing a bacterial disease, as components of antibacterial  
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
 PT plants.

XX Example; SEQ ID NO 5069; 328pp; English.

XX The invention relates to isolated *Acinetobacter baumannii* nucleic acids.  
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
 CC for diagnosing a bacterial disease, as components of antibacterial  
 CC vaccines, as targets for antibacterial drugs, to detect the presence of  
 CC A. baumannii and other *Acinetobacter* species in a sample, in screening  
 CC compounds for the ability to interfere with the A. baumannii life cycle  
 CC or to inhibit A. baumannii infection, and as biocontrol agents for  
 CC plants. The present sequence represents the amino acid sequence of an A.  
 CC baumannii protein.

XX Sequence 453 AA;

XX Query Match 10.3%; Score 96.5; DB 6; Length 453;  
 XX Best Local Similarity 29.5%; Pred. No. 0.12;  
 XX Matches 39; Conservative 18; Mismatches 42; Indels 33; Gaps 7;



Db 123 VQ-----DVAKTDPGSLSTVIAAFVNFGLVSS 152  
 QY 161 --TVEETKLAGMTDHIATSTSY 181  
 Db 153 SHNTDQDALALRLTTLTAQTATY 175

RESULT 10  
 AAR88013  
 ID AAR88013 standard, protein, 319 AA.  
 XX  
 AC AAR88013;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 02-AUG-1996 (first entry)  
 XX  
 DE Mature Pseudomonas glumae lipase T233R variant.  
 XX  
 KM Mutant; lipase; enzymatic detergent; substitution; variant; improved;  
 KM hydrophobicity.  
 OS Burkholderia glumae.  
 XX

Key Location/Qualifiers  
 FT Misc-difference 23 /note= "possible site for Phe to Arg substitution"  
 FT Misc-difference 129 /note= "possible site for Thr to Tyr substitution"  
 FT Misc-difference 134 /note= "possible site for Leu to Arg substitution"  
 FT Misc-difference 148 /note= "possible site for Thr to Val substitution"  
 FT Misc-difference 233 /label= substitution  
 FT /note= "Thr to Arg, in wild-type sequence a Thr residue is present at position 233 of mature P. glumae lipase"  
 FT Misc-difference 234 /note= "possible site for Leu to Arg substitution"  
 FT Misc-difference 239 /note= "possible site for Val to Phe substitution"  
 FT Misc-difference 240 /note= "possible site for Thr to Leu substitution"  
 FT Misc-difference 282 /note= "possible site for His to Arg substitution"  
 FT Misc-difference 292 /note= "possible site for Leu to Phe substitution"  
 FT

MO9535381-A1.  
 XX  
 PD 28-DEC-1995.  
 XX  
 PF 15-JUN-1995; 95WO-EP002349.  
 XX  
 PR 20-JUN-1994; 94EP-00201761.  
 XX  
 PA (UNITL ) UNILEVER NV.  
 PA (UNITL ) UNILEVER PLC.  
 PI Frenken IGT, Peters H, Suerbaum HW, De Vlieg J, Verrips CT;  
 DR WPI; 1996-058418/06.  
 XX  
 PT Pseudomonas lipase variant with increased surface hydrophobicity - has  
 XX improved in-the-wash performance; useful in enzymatic detergent.  
 PS  
 XX Claim 12; Page; 33pp; English.  
 CC AAR88009-R88018 are protein variants of Pseudomonas glumae lipase. The  
 CC sequences are based upon the wild-type P. glumae sequence disclosed in  
 CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of  
 CC the amino acid substitutions indicated in the features table. Using these  
 CC variants it was shown that lipases can be modified in a such a way that  
 CC interaction with the substrate can be improved without forming large

CC hydrophobic areas on the modified lipase surface which allow aggregation  
 CC of lipase molecules, the lipase variants show improved in-the-wash  
 CC lipolytic activity and may be used in enzymatic detergent compans.  
 CC (updated on 16-Oct-2003 to standardise OS field)  
 XX

SQ Sequence 319 AA;  
 QY 3 IIVHGL-----YHGLVNHPLSHRLKGYRTQTSYNSLAID-----EAI 45  
 Db 11 VILVHGLAGDKFRANVVDYWGIGSDLQSH-----GAXYVANLSGFGSDDDGNRGEQL 65  
 QY 46 FRILDRSL--THASPAIVGSHLGLGIVIRYIESRPSCEIASHVVAISPIQASIVVK 103  
 Db 66 LAVKQVLAATGATYKVLIGHSGGILT--SRVYAAVAP--QLVAVTTTGTTPRGSFPADF 122  
 QY 104 IEQLGIVALGNSAEFGLEKHDESRYPQSGSIAGTIPLGRSLILRPLSDGTV--- 160  
 Db 123 VQ-----DVAKTDPGSLSTVIAAFVNFGLVSS 152

QY 161 --TVEETKLAGMTDHIATSTSY 181  
 Db 153 SHNTDQDALALRLTTLTAQTATY 175

RESULT 11  
 AAR88015  
 ID AAR88015 standard, protein, 319 AA.  
 XX  
 AC AAR88015;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 02-AUG-1996 (first entry)  
 XX  
 DE Mature Pseudomonas glumae lipase V239F variant.  
 XX  
 KM Mutant; lipase; enzymatic detergent; substitution; variant; improved;  
 KM hydrophobicity.  
 OS Burkholderia glumae.  
 XX

Key Location/Qualifiers  
 FT Misc-difference 23 /note= "possible site for Phe to Arg substitution"  
 FT Misc-difference 129 /note= "possible site for Thr to Tyr substitution"  
 FT Misc-difference 134 /note= "possible site for Leu to Arg substitution"  
 FT Misc-difference 148 /note= "possible site for Thr to Val substitution"  
 FT Misc-difference 233 /note= "possible site for Thr to Arg substitution"  
 FT Misc-difference 234 /note= "possible site for Thr to Arg substitution"  
 FT Misc-difference 239 /note= "possible site for Leu to Arg substitution"  
 FT /label= substitution  
 FT /note= "Val to Phe, in wild-type sequence a Val residue is present at position 239 of mature P. glumae lipase"  
 FT Misc-difference 240 /note= "possible site for Thr to Leu substitution"  
 FT Misc-difference 282 /note= "possible site for His to Arg substitution"  
 FT Misc-difference 292 /note= "possible site for Leu to Phe substitution"  
 FT

MO9535381-A1.  
 XX  
 PD 28-DEC-1995.  
 XX  
 PF 15-JUN-1995; 95WO-EP002349.

XX 20-JUN-1994; 94EP-00201761.  
 PR (UNITL ) UNILEVER NV.  
 PA (UNITL ) UNILEVER PLC.  
 XX  
 PI Frenken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;  
 DR WPI; 1996-058418/06.  
 XX  
 PT Pseudomonas lipase variant with increased surface hydrophobicity - has  
 improved in-the-wash performance, useful in enzymatic detergent.  
 XX  
 PS Claim 12; Page; 33pp; English.  
 XX  
 CC AAR88009-R88018 are protein variants of Pseudomonas glumae lipase. The  
 CC sequences are based upon the wild-type P. glumae sequence disclosed in  
 CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of  
 CC the amino acid substitutions indicated in the features table. Using these  
 CC variants it was shown that lipases can be modified in a such a way that  
 CC interaction with the substrate can be improved without forming large  
 CC hydrophobic areas on the modified lipase surface which allow aggregation  
 CC of lipase molecules. The lipase variants show improved in-the-wash  
 CC lipolytic activity and may be used in enzymatic detergent compans.  
 CC (Updated on 16-Oct-2003 to standardise OS field)  
 CC  
 SQ Sequence 319 AA;  
 XX  
 Query Match 10.3%; Score 96; DB 2; Length 319;  
 Best Local Similarity 21.2%; Pred. No. 0.081;  
 Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;  
 QY 3 IIVVHGL-----YHGLVWHPLSRLHKLGYRTQTISYNSLAID-----EAI 45  
 DB 11 VILVHGLAGTDKRVANVVDYWGIGSDLOSH-----GAKYVYANLGFQSDDPNGRGHQL 65  
 QY 46 FRRLDRL--THASPNALVGHSLGGLVIRKYLESRAPSCETLSHVVAIGSPLOGASIVNK 103  
 DB 66 LAYVKQVLAATGATKTNLIGHSGQGLT--SRVYAAVAP--QVVASVTTTGTTPHSGSERADF 122  
 QY 104 IEOLGVALGNSAEFLKEHDESRYPQKSGSIAGTIPGLRSLRLRPLDSDGTV-- 160  
 DB 123 VQ-----DVLKTPDPTGLSSTVIAAFVNFGLVSS 152  
 QY 161 --TVBETKIAGMTDHAISTSY 181  
 DB 153 SHNTDDALALRLTLTTAQTATY 175  
 RESULT 12  
 AAR88018  
 ID AAR88018 standard; protein; 319 AA.  
 XX  
 AC AAR88018;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 02-AUG-1996 (first entry)  
 XX  
 DS Mature Pseudomonas glumae lipase L292F variant.  
 XX  
 KM Mutant; lipase; enzymatic detergent; substitution; variant; improved;  
 KM hydrophobicity.  
 XX  
 OS Burkholderia glumae.  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 23 /note= "possible site for Phe to Arg substitution"  
 FT Misc-difference 129 /note= "possible site for Thr to Tyr substitution"  
 FT Misc-difference 134 /note= "possible site for Leu to Arg substitution"  
 FT Misc-difference 148

FT /note= "possible site for Thr to Val substitution"  
 FT 233  
 FT Misc-difference 234 /note= "possible site for Thr to Arg substitution"  
 FT 234  
 FT Misc-difference 239 /note= "possible site for Leu to Arg substitution"  
 FT 239  
 FT Misc-difference 240 /note= "possible site for Val to Phe substitution"  
 FT 240  
 FT Misc-difference 282 /note= "possible site for Thr to Leu substitution"  
 FT 282  
 FT Misc-difference 292 /note= "possible site for His to Arg substitution"  
 FT 292  
 FT /label= substitution  
 FT /note= "Leu to Phe, in wild-type sequence a Leu residue  
 is present at position 292 of mature P. glumae lipase"  
 XX  
 XX W09535381-A1.  
 XX  
 XX 28-DEC-1995.  
 XX  
 XX 15-JUN-1995; 95MO-EP002349.  
 XX  
 XX 20-JUN-1994; 94EP-00201761.  
 XX  
 PA (UNITL ) UNILEVER NV.  
 PA (UNITL ) UNILEVER PLC.  
 XX  
 PI Frenken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;  
 DR WPI; 1996-058418/06.  
 XX  
 PT Pseudomonas lipase variant with increased surface hydrophobicity - has  
 PT improved in-the-wash performance, useful in enzymatic detergent.  
 XX  
 PS Claim 12; Page; 33pp; English.  
 XX  
 CC AAR88009-R88018 are protein variants of Pseudomonas glumae lipase. The  
 CC sequences are based upon the wild-type P. glumae sequence disclosed in  
 CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of  
 CC the amino acid substitutions indicated in the features table. Using these  
 CC variants it was shown that lipases can be modified in a such a way that  
 CC interaction with the substrate can be improved without forming large  
 CC hydrophobic areas on the modified lipase surface which allow aggregation  
 CC of lipase molecules. The lipase variants show improved in-the-wash  
 CC lipolytic activity and may be used in enzymatic detergent compans.  
 CC (Updated on 16-Oct-2003 to standardise OS field)  
 CC  
 SQ Sequence 319 AA;  
 XX  
 Query Match 10.3%; Score 96; DB 2; Length 319;  
 Best Local Similarity 21.2%; Pred. No. 0.081;  
 Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;  
 QY 3 IIVVHGL-----YHGLVWHPLSRLHKLGYRTQTISYNSLAID-----EAI 45  
 DB 11 VILVHGLAGTDKRVANVVDYWGIGSDLOSH-----GAKYVYANLGFQSDDPNGRGHQL 65  
 QY 46 FRRLDRL--THASPNALVGHSLGGLVIRKYLESRAPSCETLSHVVAIGSPLOGASIVNK 103  
 DB 66 LAYVKQVLAATGATKTNLIGHSGQGLT--SRVYAAVAP--QVVASVTTTGTTPHSGSERADF 122  
 QY 104 IEOLGVALGNSAEFLKEHDESRYPQKSGSIAGTIPGLRSLRLRPLDSDGTV-- 160  
 DB 123 VQ-----DVLKTPDPTGLSSTVIAAFVNFGLVSS 152  
 QY 161 --TVBETKIAGMTDHAISTSY 181  
 DB 153 SHNTDDALALRLTLTTAQTATY 175  
 RESULT 13  
 AAR88009  
 ID AAR88009 standard; protein; 319 AA.



```

XX AC AAR88009;
XX XX 16-OCT-2003 (revised)
XX DT 02-AUG-1996 (first entry)
XX XX Mature Pseudomonas glumae F23R variant.
XX DE Mature Pseudomonas glumae F23R variant.
XX XX Mutant; lipase; enzymatic detergent; substitution; variant; improved;
XX KM hydrophobicity.
XX OS Burkholderia glumae.
XX XX Key Location/Qualifiers
XX FH MISC-difference 23 /label= substitution
XX FT /note= "Phe to Arg, in wild-type sequence a Phe residue
XX FT is present at position 23 of mature P. glumae lipase"
XX FT MISC-difference 129 /note= "possible site for Thr to Tyr substitution"
XX FT MISC-difference 134 /note= "possible site for Leu to Arg substitution"
XX FT MISC-difference 148 /note= "possible site for Thr to Val substitution"
XX FT MISC-difference 233 /note= "possible site for Thr to Arg substitution"
XX FT MISC-difference 234 /note= "possible site for Leu to Arg substitution"
XX FT MISC-difference 239 /note= "possible site for Val to Phe substitution"
XX FT MISC-difference 240 /note= "possible site for Thr to Leu substitution"
XX FT MISC-difference 282 /note= "possible site for His to Arg substitution"
XX FT MISC-difference 292 /note= "possible site for Leu to Phe substitution"
XX FT MISC-difference 292 /note= "possible site for Leu to Phe substitution"
XX XX W09535381-A1.
XX PN 28-DEC-1995.
XX PD 15-JUN-1995; 95WO-EP002349.
XX PF 20-JUN-1994; 94EP-00201761.
XX PR (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX FT Franken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;
XX DR WPI; 1996-058418/06.
XX XX Pseudomonas lipase variant with increased surface hydrophobicity - has
XX PT improved in-the-wash performance, useful in enzymatic detergent.
XX PS Claim 12, Page; 33pp; English.
XX XX AAR88009-R88018 are protein variants of Pseudomonas glumae lipase. The
XX CC sequences are based upon the wild-type P. glumae sequence disclosed in
XX CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of
XX CC the amino acid substitutions indicated in the features table. Using these
XX CC variants it was shown that lipases can be modified in a such a way that
XX CC interaction with the substrate can be improved without forming large
XX CC hydrophobic areas on the modified lipase surface which allow aggregation
XX CC of lipase molecules. The lipase variants show improved in-the-wash
XX CC lipolytic activity and may be used in enzymatic detergent compans.
XX CC (Updated on 16-OCT-2003 to standardise OS field)
XX SO Sequence 319 AA;
Query Match 10.3%; Score 96; DB 2; Length 319;
Best Local Similarity 21.2%; Pred. No. 0.081;
Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;

```

```

QY 3 IIVVGL-----YAGLVNHPISRLAKGRTQTISVNSLAID-----EAI 45
DB 11 VILVHGLAGTDRANVVDYWGIGSDLSH-----GATVYVANSFGSDGPNRGSQL 65
QY 46 FRLLDRSL--THASPNALVGHSLGELVTKRYLESRAPSCETLSHVVAIGSPLOGASIVNK 103
DB 66 LAYVQVLAATGATATKYNLIGHGGLT--SRVYAAVAP--QVAVSYTTGTFRGSEFADF 122
QY 104 IEQLGLVVALGNSAEFGLEKHDESRYPQKSGSIAGTIPLGIRSLRLDPLDSGTV-- 160
DB 123 VQ-----DVAKTPTGTLSSTVIAAFVVFGLTVSS 152
QY 161 --TVEETKLAGTWDHIAISTTSY 181
DB 153 SHNTQDAAALRLTTLTAQTATY 175
RESULT 14
AAR88014
ID AAR88014 standard; protein; 319 AA.
XX AC AAR88014;
XX XX 16-OCT-2003 (revised)
XX DT 02-AUG-1996 (first entry)
XX DE Mature Pseudomonas glumae lipase L234R variant.
XX XX Mutant; lipase; enzymatic detergent; substitution; variant; improved;
XX KM hydrophobicity.
XX OS Burkholderia glumae.
XX XX Key Location/Qualifiers
XX FH MISC-difference 23 /note= "possible site for Phe to Arg substitution"
XX FT MISC-difference 129 /note= "possible site for Thr to Tyr substitution"
XX FT MISC-difference 134 /note= "possible site for Leu to Arg substitution"
XX FT MISC-difference 148 /note= "possible site for Thr to Val substitution"
XX FT MISC-difference 233 /note= "possible site for Thr to Arg substitution"
XX FT MISC-difference 234 /note= "possible site for Thr to Arg substitution"
XX FT /label= substitution
XX FT /note= "Leu to Arg, in wild-type sequence a Leu residue
XX FT is present at position 234 of mature P. glumae lipase"
XX FT MISC-difference 239 /note= "possible site for Val to Phe substitution"
XX FT MISC-difference 240 /note= "possible site for Thr to Leu substitution"
XX FT MISC-difference 282 /note= "possible site for His to Arg substitution"
XX FT MISC-difference 292 /note= "possible site for Leu to Phe substitution"
XX FT /note= "possible site for Leu to Phe substitution"
XX XX W09535381-A1.
XX PN 28-DEC-1995.
XX PD 15-JUN-1995; 95WO-EP002349.
XX PF 20-JUN-1994; 94EP-00201761.
XX PR (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX FT Franken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;
XX DR WPI; 1996-058418/06.

```

PT Pseudomonas lipase variant with increased surface hydrophobicity - has  
PT improved in-the-wash performance, useful in enzymatic detergent.  
XX  
PS Claim 12; Page; 33pp; English.  
CC AAR8009-R88018 are protein variants of Pseudomonas glumae lipase. The  
CC sequences are based upon the wild-type P. glumae sequence disclosed in  
CC EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of  
CC the amino acid substitutions indicated in the features table. Using these  
CC variants it was shown that lipases can be modified in a such a way that  
CC interaction with the substrate can be improved without forming large  
CC hydrophobic areas on the modified lipase surface which allow aggregation  
CC of lipase molecules. The lipase variants show improved in-the-wash  
CC lipolytic activity and may be used in enzymatic detergent compans.  
CC (Updated on 16-OCT-2003 to standardise OS field)  
SQ Sequence 319 AA;  
Query Match 10.3%; Score 96; DB 2; Length 319;  
Best Local Similarity 21.2%; Pred. No. 0.081;  
Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;  
QY 3 IILVHGL-----YHGLVHMPHSRLHKLGYRQTISYNSLAIDP-----EAI 45  
DB 11 VILVHGLAGTDKFRANVVDYVYGIQSDLOSH-----GAKYVYANLSGFQSDGPNRGEGQL 65  
QY 46 FRRLDRSL--THASPNALVGHSLGIVIKRYLESRAPSCETLSHYVAIGSPLOGASIVNK 103  
DB 66 LAVYKQVLAATGATKTNKVLIGHSGQGLT-SRYVAAVAP--QIVASVTTTGTFRHSGSEFADF 122  
QY 104 IEOLGLGVNLGNSAEFGKHEHDESRYPKSGSIAGTIPLGKRSLLRDPDSDGTV--- 160  
DB 123 VQ-----DVKLTPGTGLSSTVIAAFVNVFGLVSS 152  
QY 161 --TVEETKIAGMTDHAISTTSY 181  
DB 153 SHNTDDALAAALRTLTTAQTATY 175  
RESULT 15  
AAR8016  
ID AAR8016 standard; protein; 319 AA.  
XX  
AC AAR8016;  
DT 16-OCT-2003 (revised)  
DT 02-AUG-1996 (first entry)  
XX  
DE Mature Pseudomonas glumae lipase T240L variant.  
KM Mutant; lipase; enzymatic detergent; substitution; variant; improved;  
KM hydrophobicity.  
XX  
OS Burkholderia glumae.  
XX  
FH Key Location/Qualifiers  
FT MISC-difference 23 /note= "possible site for Phe to Arg substitution"  
FT MISC-difference 129 /note= "possible site for Thr to Tyr substitution"  
FT MISC-difference 134 /note= "possible site for Thr to Arg substitution"  
FT MISC-difference 148 /note= "possible site for Leu to Arg substitution"  
FT MISC-difference 148 /note= "possible site for Thr to Val substitution"  
FT MISC-difference 233 /note= "possible site for Thr to Arg substitution"  
FT MISC-difference 234 /note= "possible site for Thr to Arg substitution"  
FT MISC-difference 239 /note= "possible site for Leu to Arg substitution"  
FT MISC-difference 240 /note= "possible site for Val to Phe substitution"  
FT /label= substitution  
FT /note= "Thr to Leu, in wild-type sequence a Thr residue

FT is present at position 240 of mature P. glumae lipase"  
FT MISC-difference 282 /note= "possible site for His to Arg substitution"  
FT MISC-difference 292 /note= "possible site for Leu to Phe substitution"  
XX  
XX WO9535381-A1.  
XX 28-DEC-1995.  
XX  
XX 15-JUN-1995; 95WO-EP002349.  
XX 20-JUN-1994; 94EP-00201761.  
XX  
XX (UNIL ) UNILEVER NV.  
XX (UNIL ) UNILEVER PLC.  
XX  
XX Frenken LGJ, Peters H, Suerbaum HM, De Vlieg J, Verrips CT;  
XX WPI; 1996-058418/06.  
XX  
XX Pseudomonas lipase variant with increased surface hydrophobicity - has  
XX improved in-the-wash performance, useful in enzymatic detergent.  
XX  
XX Claim 12; Page; 33pp; English.  
XX  
XX AAR8009-R88018 are protein variants of Pseudomonas glumae lipase. The  
XX sequences are based upon the wild-type P. glumae sequence disclosed in  
XX EP407225-A (UNILEVER PLC). The lipase variants may contain one or more of  
XX the amino acid substitutions indicated in the features table. Using these  
XX variants it was shown that lipases can be modified in a such a way that  
XX interaction with the substrate can be improved without forming large  
XX hydrophobic areas on the modified lipase surface which allow aggregation  
XX of lipase molecules. The lipase variants show improved in-the-wash  
XX lipolytic activity and may be used in enzymatic detergent compans.  
XX (Updated on 16-OCT-2003 to standardise OS field)  
SQ Sequence 319 AA;  
Query Match 10.3%; Score 96; DB 2; Length 319;  
Best Local Similarity 21.2%; Pred. No. 0.081;  
Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;  
QY 3 IILVHGL-----YHGLVHMPHSRLHKLGYRQTISYNSLAIDP-----EAI 45  
DB 11 VILVHGLAGTDKFRANVVDYVYGIQSDLOSH-----GAKYVYANLSGFQSDGPNRGEGQL 65  
QY 46 FRRLDRSL--THASPNALVGHSLGIVIKRYLESRAPSCETLSHYVAIGSPLOGASIVNK 103  
DB 66 LAVYKQVLAATGATKTNKVLIGHSGQGLT-SRYVAAVAP--QIVASVTTTGTFRHSGSEFADF 122  
QY 104 IEOLGLGVNLGNSAEFGKHEHDESRYPKSGSIAGTIPLGKRSLLRDPDSDGTV--- 160  
DB 123 VQ-----DVKLTPGTGLSSTVIAAFVNVFGLVSS 152  
QY 161 --TVEETKIAGMTDHAISTTSY 181  
DB 153 SHNTDDALAAALRTLTTAQTATY 175  
Search completed: April 29, 2004, 11:35:16  
Job time : 60 secs













Mon May 3 09:19:52 2004

us-10-603-260-5.rpr

Page 6

Db 224 INLAOLKSVRIRREFDDLIT--ARIHGVADAI 253

Search completed: April 29, 2004, 11:37:13  
Job time : 23 secs



DR EMBL: AE004362; AAF96133.1; ALT\_INT.  
 DR PIR: S15911; S15911.  
 DR HSSP: P22088; 3LIP.  
 DR TIGR: VCA0221; -.  
 DR InterPro: IPR000073; A/b\_hydrolase.  
 DR InterPro: IPR008262; Lipase\_AS.  
 DR InterPro: IPR000379; Ser esters.  
 DR Pfam: PF00561; abhydrolase; 1.  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 KM Hydrolyase; lipid degradation; Signal, Complete proteome.  
 FT SIGNAL 1 22  
 FT CHAIN 23 312  
 FT ACT\_SITE 110 110 LACTONIZING LIPASE. (BY SIMILARITY).  
 FT ACT\_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 278 278 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 312 AA; 32995 MW; 96AA00A07A5686C CRC64;  
 Query Match 10.9%; Score 102; DB 1; Length 312;  
 Best Local Similarity 25.9%; Pred. No. 0.12;  
 Matches 49; Conservative 29; Mismatches 51; Indels 60; Gaps 10;  
 QY 3 ILVHGL-----YHGLVHPLSHRLKLG---YRQTIYSNLAIDDEAIFRRL 49  
 Db 38 ILVHGLFGFDTLAGNDYFPGI--PQS--LIRDAQYVAQVSAVNSERRGEQLAOV 92  
 QY 50 DR--SLTHASPNALVGHSLGLVIRKYLESRAPSCETLSHVAISPLQASIVKRIQL 107  
 Db 93 ESTLANTGAKVNLISHSGPRTI-RYVASVAP--DIVASVTSIGVHKGSADVLR-- 147  
 QY 108 GLGVALGNASERFGLKEHDESRYPQKSGSIAGTIPLGLRSLLI-----RDPL 154  
 Db 148 --GVI-----PSGSVSEQVAVGLTQGLVALIDLSGGKAPPOPL 185  
 QY 155 DSDGTVTVE 163  
 Db 186 ASLALITTE 194  
 RESULT 2  
 LIP\_PSEPR STANDARD; PRT; 277 AA.  
 AC P08658;  
 DT 01-UN-1988 (Rel. 06, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).  
 OS Pseudomonas fragi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 ON NCBI\_TaxID=296;  
 RX STRAIN=FROM N.A.  
 RC STRAIN=IFO 12049;  
 RC MEDLINE=89078617; PubMed=3060375;  
 RA Aoyama S., Yoshida N., Inouye S.;  
 RT "Cloning, sequencing and expression of the lipase gene from  
 RT Pseudomonas fragi IFO-12049 in E. coli.";  
 RL FEBS Lett. 242:36-40(1988).  
 RN [2]  
 RP SEQUENCE OF 1-118 FROM N.A.  
 RC STRAIN=IFO 3458;  
 RC MEDLINE=87100109; PubMed=3800995;  
 RA Kugimiyu W., Otsu Y., Hashimoto Y., Takagi Y.;  
 RT "Molecular cloning and nucleotide sequence of the lipase gene from  
 RT Pseudomonas fragi.";  
 RL Biochem. Biophys. Res. Commun. 141:185-190(1986).  
 RL -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
 CC fatty acid anion.  
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS  
 CC LIPASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: M14604; AAA25879.1; ALT\_SEQ.  
 DR EMBL: X14033; CAA32193.1; -.  
 DR PIR: S02005; S02005.  
 DR HSSP: Q05489; 17AH.  
 DR InterPro: IPR000073; A/b\_hydrolase.  
 DR InterPro: IPR008262; Lipase\_AS.  
 DR InterPro: IPR000379; Ser esters.  
 DR Pfam: PF00561; abhydrolase; 1.  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 KM Hydrolyase; lipid degradation; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 277  
 FT ACT\_SITE 83 83 LIPASE.  
 FT ACT\_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 260 260 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 277 AA; 23985 MW; 9A1562B0772C997 CRC64;  
 Query Match 10.4%; Score 97; DB 1; Length 277;  
 Best Local Similarity 24.6%; Pred. No. 0.27;  
 Matches 51; Conservative 31; Mismatches 79; Indels 46; Gaps 11;  
 QY 3 ILVHGL-----YHGLVHPLSHRLKLG---YRQTIYSNLAIDDEAIFRRL 48  
 Db 11 ILVHGLFGFDRLGSHHYFPGI--IKQALNECG--ASVFPPLISANDNEAGDDQLIKQ 64  
 QY 49 LD--RSLTASPNALVGHSLGLVIRKYLESRAPSCETLSHVAISPLQASIVKRIE- 105  
 Db 65 IHNLRQVGAQRVNLIGHSGALT-ARYVAALAP--ELIASVTSVSGPNHGSRLADRLRL 121  
 QY 106 -----QLGVALGNASERF--LKEHDESRYPQKSGSIAGTIPLGLRSLLIARDPLSDS 158  
 Db 122 AFVPRGLGETVAALTTSPFAFLSALSGHPRPONA-----LNLV--NALTTDS 168  
 QY 159 TVTVEETKIAGMTDHTAISTSYENAV 185  
 Db 169 VAAFNRPQGLPDRMGCMGPAQVNAV 195  
 RESULT 3  
 LIP\_PSEGL STANDARD; PRT; 358 AA.  
 AC Q05489;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).  
 GN LIPA.  
 OS Pseudomonas glumae, and  
 OS Chromobacterium viscosum.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia.  
 ON NCBI\_TaxID=337; 42739;  
 RX [1]  
 RP SEQUENCE FROM N.A.; SEQUENCE OF 40-61, AND MUTAGENESIS.  
 RC SPECIES=P. glumae; STRAIN=PGL / CBS 322.89;  
 RX MEDLINE=93119130; PubMed=1476423;  
 RA Frenken L.G.J., Edmond M.R., Batendurg A.M., Bos J.W., Visser C.,  
 RA Verrips C.T.;  
 RT "Cloning of the Pseudomonas glumae lipase gene and determination of  
 RT the active site residues.";  
 RL Appl. Environ. Microbiol. 58:3787-3791(1992).  
 RN [2]  
 RP SEQUENCE OF 40-54, AND CHARACTERIZATION.  
 RC SPECIES=C. viscosum;  
 RX MEDLINE=85306500; PubMed=7786905;  
 RA Taira M.A., Liebeton K., Costa J.V., Cabral J.M.S., Jaeger K.-E.;  
 RA "Lipase from Chromobacterium viscosum: biochemical characterization

RT indicating homology to the lipase from *Pseudomonas glumae*.";  
 RL Biochim. Biophys. Acta 1256:396-402(1995).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
 RC SPECIES=P. glumae;  
 RX MEDLINE=94009622; PubMed=8405390;  
 RA Noble M.E.M., Cleasby A., Johnson L.N., Edmond M.R., Frenken L.G.J.;  
 RT "The crystal structure of triacylglycerol lipase from *Pseudomonas*  
 RL *glumae* reveals a partially redundant catalytic aspartate.";  
 FEBS Lett. 331:123-128(1993).  
 [4]  
 RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RP SPECIES=C. viscosum; STRAIN=ATCC 6918;  
 RX MEDLINE=96275656; PubMed=8683577;  
 RA Lang D., Hofmann B., Haalok L., Hecht H.-J., Spener F., Schmid R.D.,  
 RA Schomburg D.;  
 RT "Crystal structure of a bacterial lipase from *Chromobacterium*  
 RL *viscosum* ATCC 6918 refined at 1.6-A resolution.";  
 J. Mol. Biol. 259:704-717(1996).  
 CC -1- FUNCTION: HYDROLYSIS OF TRIGLYCERIDES.  
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
 CC fatty acid anion.  
 CC -1- COFACTOR: Requires calcium.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS  
 CC LIPASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X70354; CAA49812.1; -;  
 DR EMBL: A16323; CAA01279.1; -;  
 DR EMBL: A32021; CAA02073.1; -;  
 DR PIR: A48952; A48952.  
 DR PDB: 1TAH; 31-MAY-94.  
 DR PDB: 1CVL; 01-APR-97.  
 DR PDB: 1QGE; 06-MAY-99.  
 DR InterPro: IPR000073; A/b\_hydrolase.  
 DR InterPro: IPR008262; Lipase\_AS.  
 DR InterPro: IPR00379; Ser\_estrs.  
 DR Pfam: PF00561; abhydrolase; 1.  
 DR ProSite: PS00120; LIPASE\_SER; 1.  
 KW Hydrolyase; Lipid degradation; Signal; Calcium; 3D-structure.  
 FT SIGNAL 1 39  
 FT CHAIN 40 358  
 FT ACT\_SITE 126 126 LIPASE.  
 FT ACT\_SITE 302 302 CHANGE RELAY SYSTEM.  
 FT ACT\_SITE 324 324 CHANGE RELAY SYSTEM.  
 FT DISULFID 229 308 CHANGE RELAY SYSTEM.  
 FT MUTAGEN 54 54 H->A: NO LOSS OF ACTIVITY.  
 FT MUTAGEN 126 126 S->A: COMPLETE LOSS OF ACTIVITY.  
 FT MUTAGEN 160 160 D->E: NO LOSS OF ACTIVITY.  
 FT MUTAGEN 160 160 D->A: NO LOSS OF ACTIVITY.  
 FT MUTAGEN 280 280 D->E: NO LOSS OF ACTIVITY.  
 FT MUTAGEN 280 280 D->A: COMPLETE LOSS OF ACTIVITY.  
 FT MUTAGEN 302 302 D->E: NO LOSS OF ACTIVITY.  
 FT MUTAGEN 302 302 D->A: 75% LOSS OF ACTIVITY.  
 FT MUTAGEN 324 324 H->A: COMPLETE LOSS OF ACTIVITY.  
 FT CONFLICT 40 40 A -> W (IN REF. 2).  
 FT TURN 43 44  
 FT STRAND 50 53  
 FT TURN 61 62  
 FT HELIX 65 67  
 FT TURN 70 71  
 FT HELIX 72 78  
 FT TURN 79 80  
 FT STRAND 83 86

FT TURN 89 90  
 FT TURN 97 98  
 FT HELIX 100 115  
 FT TURN 116 116  
 FT STRAND 120 125  
 FT TURN 126 126  
 FT HELIX 127 138  
 FT HELIX 140 142  
 FT STRAND 143 149  
 FT TURN 153 154  
 FT HELIX 157 168  
 FT TURN 170 171  
 FT TURN 173 174  
 FT HELIX 176 188  
 FT TURN 191 193  
 FT HELIX 196 204  
 FT TURN 205 206  
 FT HELIX 208 217  
 FT TURN 221 222  
 FT STRAND 223 223  
 FT TURN 226 228  
 FT STRAND 231 231  
 FT STRAND 235 238  
 FT TURN 239 240  
 FT STRAND 241 248  
 FT STRAND 250 250  
 FT STRAND 253 259  
 FT TURN 260 261  
 FT STRAND 262 267  
 FT TURN 275 275  
 FT HELIX 276 279  
 FT HELIX 281 294  
 FT TURN 295 297  
 FT STRAND 301 301  
 FT STRAND 306 306  
 FT HELIX 307 310  
 FT STRAND 314 315  
 FT STRAND 320 320  
 FT TURN 324 325  
 FT HELIX 326 328  
 FT TURN 329 333  
 FT TURN 337 338  
 FT HELIX 342 356  
 FT TURN 357 358  
 SQ SEQUENCE 358 AA; 36928 MW; FE7B5D7A22EC6B4B CRC64;  
 Query Match 10.3%; Score 96; DB 1; Length 358;  
 Best Local Similarity 21.2%; Pred. No. 0.45;  
 Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;  
 QY 3 IIVHGL-----YHGLVHPLSHRLHLGRTQTITISNLAIID-----EAI 45  
 DB 50 VILVHGLAGTDKPFANVVDVWYIGISDLQSH-----GAKVYVANTSGFQSDGPNRGHQQL 104  
 QY 46 FRLDRSL--THASPNALVHSLGLVTKRYLESAPFCETLSHVVAIGSPLOGASIVNK 103  
 DB 105 IAVVQVLAATGATATNVLNIGSGGLT-SRYVAAP--QVAVSTTTGTFRHGEFFVDF 161  
 QY 104 IEQLGLVALGNSAEFGKHEHDESRYPQKSGSIAGTIPILGRLSLRDPPLDSGTV-- 160  
 DB 162 VQ-----DVAKTPTGISSTVIAFAVNVFGLVSS 191  
 QY 161 --TVEETKIAGMTDHIATISYSY 181  
 DB 192 SHNTDQDALMRITLTAAQTATY 214  
 RESULT 4  
 YHET ECOLI  
 ID YHET ECOLI STANDARD; PRT; 340 AA.  
 AC P45524;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DN Hypothetical protein yhet.  
 GN Yhet OR B3353.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxId=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shoeh Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474 (1997).  
 CC -1- SIMILARITY: Belongs to the UPF0017 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U18997; AAA58150.1; -.  
 DR EMBL; AE000411; AAC76376.1; -.  
 DR PIR; D65128; D65129.  
 DR Ecogene; E612904; yhet.  
 DR InterPro; IPR000073; A/b\_hydrolase.  
 DR InterPro; IPR000379; Ser\_ester.  
 DR InterPro; IPR000952; UPF0017.  
 DR Pfam; PF00561; abhydrolase\_1.  
 DR PROSITE; PS01133; UPF0017; 1.  
 DR Hypothetical protein Transmembrane; Complete proteome.  
 KM TRANSMEM 147 164  
 FT POTENTIAL.  
 SQ SEQUENCE 340 AA; 38495 MW; 7DB6BA6795CBA64D CRC64;  
 Query Match 10.2%; Score 95.5; DB 1; Length 340;  
 Best Local Similarity 24.9%; Pred. No. 0.46;  
 Matches 53; Conservative 27; Mismatches 60; Indels 73; Gaps 13;  
 QY 2 QILVHGL-----YMHGL-----VME-----PUSHRLHKTGYRTQTSY 35  
 DB 74 RLAVFEGLEGLSINSPYAHGLVAAQKRGWLVYMHFRGSGSEP--NRHRI-----Y 123  
 QY 36 NSLAIDDEAIPRR-LDRSLTHASPNALVGHSLGIVIKRYLESRAPSCETLSHVVALGSP 94  
 DB 124 HSGETEDASWFLRMLOKRGHA-PTNAVGYSLGNNMLACLAKENDL-----P 171  
 QY 95 LOGASIVN--KIEOLGVALGNSAEFGLEKHEHDESR-----PQSGSIAGTIP 142  
 DB 172 VDAAVIVSAPFMLEACSYHEKGSFRVY-----QRYLNLKANAARKLAAYPGTLP 223  
 QY 143 LGLRSL-LLRDPLDSGTYVTEETKAGTTHI 174  
 DB 224 INLAQKSVRIREFDILT--ARLHGADAI 253  
 RESULT 5  
 LIP\_PSS5 STANDARD; PRT; 364 AA.  
 AC F25275;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase).  
 GN LIP.  
 OS Pseudomonas sp. (strain KMI-56).  
 OC Bacteria; Proteobacteria.  
 OX NCBI\_TaxId=311;

RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 45-47.  
 RX MEDLINE=92118328; PubMed=1358739;  
 RA Iizumi T., Nakamura K., Shimada Y., Sugihara A., Tomimaga Y.,  
 RA Fukase T.;  
 RT "Cloning, nucleotide sequencing, and expression in Escherichia coli  
 RT of a lipase and its activator genes from Pseudomonas sp. KMI-56.";  
 RL Agric. Biol. Chem. 55:2349-2357 (1991).  
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H<sub>2</sub>O = diacylglycerol + a  
 CC fatty acid anion.  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE SUPERFAMILY. PSEUDOMONAS  
 CC LIPASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D10069; BAA00960.1; -.  
 DR EMBL; S77842; AAC60400.1; -.  
 DR HSP; P22088; 3LIP.  
 DR InterPro; IPR000073; A/b\_hydrolase.  
 DR InterPro; IPR008262; Lipase\_AS.  
 DR InterPro; IPR000379; Ser\_ester.  
 DR Pfam; PF00561; abhydrolase\_1.  
 DR PROSITE; PS00120; LIPASE\_SER; 1.  
 DR Hydrolase; Lipid degradation; Signal.  
 KM SIGNAL 1 44  
 FT CHAIN 45 364  
 FT ACT\_SITE 131 131  
 FT ACT\_SITE 286 286  
 FT ACT\_SITE 330 330  
 SQ SEQUENCE 364 AA; 37511 MW; F34C8B82B94E27D CRC64;  
 Query Match 9.7%; Score 90.5; DB 1; Length 364;  
 Best Local Similarity 26.6%; Pred. No. 1.4;  
 Matches 33; Conservative 21; Mismatches 43; Indels 27; Gaps 5;  
 QY 3 IILVHGL-----YMHGLVMEPLSHRLHKTGYRTQTSYNSLAIDDEA----- 44  
 DB 55 IILVHGLSGDKKAGVETWYG-----IQEDLQGNATVYVNLSEFGDDGANGREGOL 109  
 QY 45 -IFRLDRSLTHASPNALVGHSLGIVIKRYLESRAPSCETLSHVVALGSP 103  
 DB 110 LAVVKVLAATGATKVLVGHSGGLT-SRYVAAP--DLVASTTIGTPHRSSEFADF 166  
 QY 104 IEOL 107  
 DB 167 VQNV 170  
 RESULT 6  
 ATC6\_YEAST STANDARD; PRT; 1215 AA.  
 ID ATC6\_YEAST  
 AC P39986;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable cation-transporting ATPase 1 (EC 3.6.3.-).  
 GN SP1 OR YEL031W.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313264; PubMed=9169868;  
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Yelton M.A., Allen E.,  
 RA Araujo R., Aviles E., Bernan A., Brennan T., Carpenter J., Chen E.,

RA Cheray J.M., Chung E., Duncan M., Guzman E., Hartzell G.,  
 RA Huncke-Smith S., Hyman R.W., Kayser A., Komp C., Laabkari D., Lew H.,  
 RA Lin D., Moseedale D., Nakahara K., Namath A., Norgren R., Oefner P.,  
 RA Oh C., Peck F.X., Roberts D., Sehl P., Schramm S., Shogren T.,  
 RA Smith V., Taylor P., Wei Y., Weinstein D., Davis R.W.,  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.",  
 RN Nature 387:78-81(1997).  
 CC [2]  
 CC CHARACTERIZATION.  
 CC MEDLINE=99291048; PubMed=10361284;  
 RA Suzuki C., Shima Y.I.;  
 RT "P-type ATPase spf1 mutants show a novel resistance mechanism for the  
 RT killer toxin SMK1".  
 CC Mol. Microbiol. 32:813-823(1999).  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: Belongs to the cation transport ATPases family (P-type  
 CC ATPases). Subfamily V.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, U18530; AAB64508.1; -.  
 CC PIR, S50428; S50428.  
 CC Germonline, 139035; -.  
 CC SGD, S0000757; SPFL.  
 CC GO, GO:0005789; C:cytoplasmic reticulum membrane, IDA.  
 CC GO, GO:0015662; F:ATPase activity, coupled to transmembrane m. . .; IDA.  
 CC GO, GO:0006874; P:calcium ion homeostasis; IMP.  
 CC GO, GO:0006486; P:protein amino acid glycosylation; IMP.  
 CC InterPro, IPR001757; ATPase\_E1-E2.  
 CC InterPro, IPR008250; E1-E2\_ATPase\_reg.  
 CC InterPro, IPR005834; Hydrolyase.  
 CC InterPro, IPR006544; P-ATPase-V.  
 CC Pfam, PF00122; E1-E2\_ATPase; 1.  
 CC Pfam, PF00702; Hydrolyase; 1.  
 CC PRINTS, PRO0119; CATATPASE.  
 CC TIGRFS, TIGR01494; ATPase\_P-type; 6.  
 CC TIGRFS, TIGR01657; P-ATPase-V; 1.  
 CC PROSITE, PS00554; ATPase\_E1-E2; 1.  
 KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.  
 FT DOMAIN 1 21  
 FT TRANSSEM 22 43  
 FT TRANSSEM 44 49  
 FT TRANSSEM 50 72  
 FT TRANSSEM 73 191  
 FT TRANSSEM 192 214  
 FT TRANSSEM 215 217  
 FT TRANSSEM 218 236  
 FT TRANSSEM 237 399  
 FT TRANSSEM 400 419  
 FT TRANSSEM 420 432  
 FT TRANSSEM 433 454  
 FT TRANSSEM 455 493  
 FT TRANSSEM 994 1013  
 FT TRANSSEM 1014 1020  
 FT TRANSSEM 1021 1037  
 FT TRANSSEM 1038 1055  
 FT TRANSSEM 1056 1079  
 FT TRANSSEM 1080 1099  
 FT TRANSSEM 1100 1122  
 FT TRANSSEM 1123 1133  
 FT TRANSSEM 1134 1153  
 FT TRANSSEM 1154 1170  
 FT TRANSSEM 1171 1193  
 FT TRANSSEM 1194 1215  
 FT TRANSSEM 1215 1287  
 FT TRANSSEM 1287 1487  
 FT TRANSSEM 1487 1616  
 FT TRANSSEM 1616 1816  
 FT TRANSSEM 1816 2016  
 FT TRANSSEM 2016 2216  
 FT TRANSSEM 2216 2416  
 FT TRANSSEM 2416 2616  
 FT TRANSSEM 2616 2816  
 FT TRANSSEM 2816 3016  
 FT TRANSSEM 3016 3216  
 FT TRANSSEM 3216 3416  
 FT TRANSSEM 3416 3616  
 FT TRANSSEM 3616 3816  
 FT TRANSSEM 3816 4016  
 FT TRANSSEM 4016 4216  
 FT TRANSSEM 4216 4416  
 FT TRANSSEM 4416 4616  
 FT TRANSSEM 4616 4816  
 FT TRANSSEM 4816 5016  
 FT TRANSSEM 5016 5216  
 FT TRANSSEM 5216 5416  
 FT TRANSSEM 5416 5616  
 FT TRANSSEM 5616 5816  
 FT TRANSSEM 5816 6016  
 FT TRANSSEM 6016 6216  
 FT TRANSSEM 6216 6416  
 FT TRANSSEM 6416 6616  
 FT TRANSSEM 6616 6816  
 FT TRANSSEM 6816 7016  
 FT TRANSSEM 7016 7216  
 FT TRANSSEM 7216 7416  
 FT TRANSSEM 7416 7616  
 FT TRANSSEM 7616 7816  
 FT TRANSSEM 7816 8016  
 FT TRANSSEM 8016 8216  
 FT TRANSSEM 8216 8416  
 FT TRANSSEM 8416 8616  
 FT TRANSSEM 8616 8816  
 FT TRANSSEM 8816 9016  
 FT TRANSSEM 9016 9216  
 FT TRANSSEM 9216 9416  
 FT TRANSSEM 9416 9616  
 FT TRANSSEM 9616 9816  
 FT TRANSSEM 9816 10016  
 FT TRANSSEM 10016 10216  
 FT TRANSSEM 10216 10416  
 FT TRANSSEM 10416 10616  
 FT TRANSSEM 10616 10816  
 FT TRANSSEM 10816 11016  
 FT TRANSSEM 11016 11216  
 FT TRANSSEM 11216 11416  
 FT TRANSSEM 11416 11616  
 FT TRANSSEM 11616 11816  
 FT TRANSSEM 11816 12016  
 FT TRANSSEM 12016 12216  
 FT TRANSSEM 12216 12416  
 FT TRANSSEM 12416 12616  
 FT TRANSSEM 12616 12816  
 FT TRANSSEM 12816 13016  
 FT TRANSSEM 13016 13216  
 FT TRANSSEM 13216 13416  
 FT TRANSSEM 13416 13616  
 FT TRANSSEM 13616 13816  
 FT TRANSSEM 13816 14016  
 FT TRANSSEM 14016 14216  
 FT TRANSSEM 14216 14416  
 FT TRANSSEM 14416 14616  
 FT TRANSSEM 14616 14816  
 FT TRANSSEM 14816 15016  
 FT TRANSSEM 15016 15216  
 FT TRANSSEM 15216 15416  
 FT TRANSSEM 15416 15616  
 FT TRANSSEM 15616 15816  
 FT TRANSSEM 15816 16016  
 FT TRANSSEM 16016 16216  
 FT TRANSSEM 16216 16416  
 FT TRANSSEM 16416 16616  
 FT TRANSSEM 16616 16816  
 FT TRANSSEM 16816 17016  
 FT TRANSSEM 17016 17216  
 FT TRANSSEM 17216 17416  
 FT TRANSSEM 17416 17616  
 FT TRANSSEM 17616 17816  
 FT TRANSSEM 17816 18016  
 FT TRANSSEM 18016 18216  
 FT TRANSSEM 18216 18416  
 FT TRANSSEM 18416 18616  
 FT TRANSSEM 18616 18816  
 FT TRANSSEM 18816 19016  
 FT TRANSSEM 19016 19216  
 FT TRANSSEM 19216 19416  
 FT TRANSSEM 19416 19616  
 FT TRANSSEM 19616 19816  
 FT TRANSSEM 19816 20016  
 FT TRANSSEM 20016 20216  
 FT TRANSSEM 20216 20416  
 FT TRANSSEM 20416 20616  
 FT TRANSSEM 20616 20816  
 FT TRANSSEM 20816 21016  
 FT TRANSSEM 21016 21216  
 FT TRANSSEM 21216 21416  
 FT TRANSSEM 21416 21616  
 FT TRANSSEM 21616 21816  
 FT TRANSSEM 21816 22016  
 FT TRANSSEM 22016 22216  
 FT TRANSSEM 22216 22416  
 FT TRANSSEM 22416 22616  
 FT TRANSSEM 22616 22816  
 FT TRANSSEM 22816 23016  
 FT TRANSSEM 23016 23216  
 FT TRANSSEM 23216 23416  
 FT TRANSSEM 23416 23616  
 FT TRANSSEM 23616 23816  
 FT TRANSSEM 23816 24016  
 FT TRANSSEM 24016 24216  
 FT TRANSSEM 24216 24416  
 FT TRANSSEM 24416 24616  
 FT TRANSSEM 24616 24816  
 FT TRANSSEM 24816 25016  
 FT TRANSSEM 25016 25216  
 FT TRANSSEM 25216 25416  
 FT TRANSSEM 25416 25616  
 FT TRANSSEM 25616 25816  
 FT TRANSSEM 25816 26016  
 FT TRANSSEM 26016 26216  
 FT TRANSSEM 26216 26416  
 FT TRANSSEM 26416 26616  
 FT TRANSSEM 26616 26816  
 FT TRANSSEM 26816 27016  
 FT TRANSSEM 27016 27216  
 FT TRANSSEM 27216 27416  
 FT TRANSSEM 27416 27616  
 FT TRANSSEM 27616 27816  
 FT TRANSSEM 27816 28016  
 FT TRANSSEM 28016 28216  
 FT TRANSSEM 28216 28416  
 FT TRANSSEM 28416 28616  
 FT TRANSSEM 28616 28816  
 FT TRANSSEM 28816 29016  
 FT TRANSSEM 29016 29216  
 FT TRANSSEM 29216 29416  
 FT TRANSSEM 29416 29616  
 FT TRANSSEM 29616 29816  
 FT TRANSSEM 29816 30016  
 FT TRANSSEM 30016 30216  
 FT TRANSSEM 30216 30416  
 FT TRANSSEM 30416 30616  
 FT TRANSSEM 30616 30816  
 FT TRANSSEM 30816 31016  
 FT TRANSSEM 31016 31216  
 FT TRANSSEM 31216 31416  
 FT TRANSSEM 31416 31616  
 FT TRANSSEM 31616 31816  
 FT TRANSSEM 31816 32016  
 FT TRANSSEM 32016 32216  
 FT TRANSSEM 32216 32416  
 FT TRANSSEM 32416 32616  
 FT TRANSSEM 32616 32816  
 FT TRANSSEM 32816 33016  
 FT TRANSSEM 33016 33216  
 FT TRANSSEM 33216 33416  
 FT TRANSSEM 33416 33616  
 FT TRANSSEM 33616 33816  
 FT TRANSSEM 33816 34016  
 FT TRANSSEM 34016 34216  
 FT TRANSSEM 34216 34416  
 FT TRANSSEM 34416 34616  
 FT TRANSSEM 34616 34816  
 FT TRANSSEM 34816 35016  
 FT TRANSSEM 35016 35216  
 FT TRANSSEM 35216 35416  
 FT TRANSSEM 35416 35616  
 FT TRANSSEM 35616 35816  
 FT TRANSSEM 35816 36016  
 FT TRANSSEM 36016 36216  
 FT TRANSSEM 36216 36416  
 FT TRANSSEM 36416 36616  
 FT TRANSSEM 36616 36816  
 FT TRANSSEM 36816 37016  
 FT TRANSSEM 37016 37216  
 FT TRANSSEM 37216 37416  
 FT TRANSSEM 37416 37616  
 FT TRANSSEM 37616 37816  
 FT TRANSSEM 37816 38016  
 FT TRANSSEM 38016 38216  
 FT TRANSSEM 38216 38416  
 FT TRANSSEM 38416 38616  
 FT TRANSSEM 38616 38816  
 FT TRANSSEM 38816 39016  
 FT TRANSSEM 39016 39216  
 FT TRANSSEM 39216 39416  
 FT TRANSSEM 39416 39616  
 FT TRANSSEM 39616 39816  
 FT TRANSSEM 39816 40016  
 FT TRANSSEM 40016 40216  
 FT TRANSSEM 40216 40416  
 FT TRANSSEM 40416 40616  
 FT TRANSSEM 40616 40816  
 FT TRANSSEM 40816 41016  
 FT TRANSSEM 41016 41216  
 FT TRANSSEM 41216 41416  
 FT TRANSSEM 41416 41616  
 FT TRANSSEM 41616 41816  
 FT TRANSSEM 41816 42016  
 FT TRANSSEM 42016 42216  
 FT TRANSSEM 42216 42416  
 FT TRANSSEM 42416 42616  
 FT TRANSSEM 42616 42816  
 FT TRANSSEM 42816 43016  
 FT TRANSSEM 43016 43216  
 FT TRANSSEM 43216 43416  
 FT TRANSSEM 43416 43616  
 FT TRANSSEM 43616 43816  
 FT TRANSSEM 43816 44016  
 FT TRANSSEM 44016 44216  
 FT TRANSSEM 44216 44416  
 FT TRANSSEM 44416 44616  
 FT TRANSSEM 44616 44816  
 FT TRANSSEM 44816 45016  
 FT TRANSSEM 45016 45216  
 FT TRANSSEM 45216 45416  
 FT TRANSSEM 45416 45616  
 FT TRANSSEM 45616 45816  
 FT TRANSSEM 45816 46016  
 FT TRANSSEM 46016 46216  
 FT TRANSSEM 46216 46416  
 FT TRANSSEM 46416 46616  
 FT TRANSSEM 46616 46816  
 FT TRANSSEM 46816 47016  
 FT TRANSSEM 47016 47216  
 FT TRANSSEM 47216 47416  
 FT TRANSSEM 47416 47616  
 FT TRANSSEM 47616 47816  
 FT TRANSSEM 47816 48016  
 FT TRANSSEM 48016 48216  
 FT TRANSSEM 48216 48416  
 FT TRANSSEM 48416 48616  
 FT TRANSSEM 48616 48816  
 FT TRANSSEM 48816 49016  
 FT TRANSSEM 49016 49216  
 FT TRANSSEM 49216 49416  
 FT TRANSSEM 49416 49616  
 FT TRANSSEM 49616 49816  
 FT TRANSSEM 49816 50016  
 FT TRANSSEM 50016 50216  
 FT TRANSSEM 50216 50416  
 FT TRANSSEM 50416 50616  
 FT TRANSSEM 50616 50816  
 FT TRANSSEM 50816 51016  
 FT TRANSSEM 51016 51216  
 FT TRANSSEM 51216 51416  
 FT TRANSSEM 51416 51616  
 FT TRANSSEM 51616 51816  
 FT TRANSSEM 51816 52016  
 FT TRANSSEM 52016 52216  
 FT TRANSSEM 52216 52416  
 FT TRANSSEM 52416 52616  
 FT TRANSSEM 52616 52816  
 FT TRANSSEM 52816 53016  
 FT TRANSSEM 53016 53216  
 FT TRANSSEM 53216 53416  
 FT TRANSSEM 53416 53616  
 FT TRANSSEM 53616 53816  
 FT TRANSSEM 53816 54016  
 FT TRANSSEM 54016 54216  
 FT TRANSSEM 54216 54416  
 FT TRANSSEM 54416 54616  
 FT TRANSSEM 54616 54816  
 FT TRANSSEM 54816 55016  
 FT TRANSSEM 55016 55216  
 FT TRANSSEM 55216 55416  
 FT TRANSSEM 55416 55616  
 FT TRANSSEM 55616 55816  
 FT TRANSSEM 55816 56016  
 FT TRANSSEM 56016 56216  
 FT TRANSSEM 56216 56416  
 FT TRANSSEM 56416 56616  
 FT TRANSSEM 56616 56816  
 FT TRANSSEM 56816 57016  
 FT TRANSSEM 57016 57216  
 FT TRANSSEM 57216 57416  
 FT TRANSSEM 57416 57616  
 FT TRANSSEM 57616 57816  
 FT TRANSSEM 57816 58016  
 FT TRANSSEM 58016 58216  
 FT TRANSSEM 58216 58416  
 FT TRANSSEM 58416 58616  
 FT TRANSSEM 58616 58816  
 FT TRANSSEM 58816 59016  
 FT TRANSSEM 59016 59216  
 FT TRANSSEM 59216 59416  
 FT TRANSSEM 59416 59616  
 FT TRANSSEM 59616 59816  
 FT TRANSSEM 59816 60016  
 FT TRANSSEM 60016 60216  
 FT TRANSSEM 60216 60416  
 FT TRANSSEM 60416 60616  
 FT TRANSSEM 60616 60816  
 FT TRANSSEM 60816 61016  
 FT TRANSSEM 61016 61216  
 FT TRANSSEM 61216 61416  
 FT TRANSSEM 61416 61616  
 FT TRANSSEM 61616 61816  
 FT TRANSSEM 61816 62016  
 FT TRANSSEM 62016 62216  
 FT TRANSSEM 62216 62416  
 FT TRANSSEM 62416 62616  
 FT TRANSSEM 62616 62816  
 FT TRANSSEM 62816 63016  
 FT TRANSSEM 63016 63216  
 FT TRANSSEM 63216 63416  
 FT TRANSSEM 63416 63616  
 FT TRANSSEM 63616 63816  
 FT TRANSSEM 63816 64016  
 FT TRANSSEM 64016 64216  
 FT TRANSSEM 64216 64416  
 FT TRANSSEM 64416 64616  
 FT TRANSSEM 64616 64816  
 FT TRANSSEM 64816 65016  
 FT TRANSSEM 65016 65216  
 FT TRANSSEM 65216 65416  
 FT TRANSSEM 65416 65616  
 FT TRANSSEM 65616 65816  
 FT TRANSSEM 65816 66016  
 FT TRANSSEM 66016 66216  
 FT TRANSSEM 66216 66416  
 FT TRANSSEM 66416 66616  
 FT TRANSSEM 66616 66816  
 FT TRANSSEM 66816 67016  
 FT TRANSSEM 67016 67216  
 FT TRANSSEM 67216 67416  
 FT TRANSSEM 67416 67616  
 FT TRANSSEM 67616 67816  
 FT TRANSSEM 67816 68016  
 FT TRANSSEM 68016 68216  
 FT TRANSSEM 68216 68416  
 FT TRANSSEM 68416 68616  
 FT TRANSSEM 68616 68816  
 FT TRANSSEM 68816 69016  
 FT TRANSSEM 69016 69216  
 FT TRANSSEM 69216 69416  
 FT TRANSSEM 69416 69616  
 FT TRANSSEM 69616 69816  
 FT TRANSSEM 69816 70016  
 FT TRANSSEM 70016 70216  
 FT TRANSSEM 70216 70416  
 FT TRANSSEM 70416 70616  
 FT TRANSSEM 70616 70816  
 FT TRANSSEM 70816 71016  
 FT TRANSSEM 71016 71216  
 FT TRANSSEM 71216 71416  
 FT TRANSSEM 71416 71616  
 FT TRANSSEM 71616 71816  
 FT TRANSSEM 71816 72016  
 FT TRANSSEM 72016 72216  
 FT TRANSSEM 72216 72416  
 FT TRANSSEM 72416 72616  
 FT TRANSSEM 72616 72816  
 FT TRANSSEM 72816 73016  
 FT TRANSSEM 73016 73216  
 FT TRANSSEM 73216 73416  
 FT TRANSSEM 73416 73616  
 FT TRANSSEM 73616 73816  
 FT TRANSSEM 73816 74016  
 FT TRANSSEM 74016 74216  
 FT TRANSSEM 74216 74416  
 FT TRANSSEM 74416 74616  
 FT TRANSSEM 74616 74816  
 FT TRANSSEM 74816 75016  
 FT TRANSSEM 75016 75216  
 FT TRANSSEM 75216 75416  
 FT TRANSSEM 75416 75616  
 FT TRANSSEM 75616 75816  
 FT TRANSSEM 75816 76016  
 FT TRANSSEM 76016 76216  
 FT TRANSSEM 76216 76416  
 FT TRANSSEM 76416 76616  
 FT TRANSSEM 76616 76816  
 FT TRANSSEM 76816 77016  
 FT TRANSSEM 77016 77216  
 FT TRANSSEM 77216 77416  
 FT TRANSSEM 77416 77616  
 FT TRANSSEM 77616 77816  
 FT TRANSSEM 77816 78016  
 FT TRANSSEM 78016 78216  
 FT TRANSSEM 78216 78416  
 FT TRANSSEM 78416 78616  
 FT TRANSSEM 78616 78816  
 FT TRANSSEM 78816 79016  
 FT TRANSSEM 79016 79216  
 FT TRANSSEM 79216 79416  
 FT TRANSSEM 79416 79616  
 FT TRANSSEM 79616 79816  
 FT TRANSSEM 79816 80016  
 FT TRANSSEM 80016 80216  
 FT TRANSSEM 80216 80416  
 FT TRANSSEM 80416 80616  
 FT TRANSSEM 80616 80816  
 FT TRANSSEM 80816 81016  
 FT TRANSSEM 81016 81216  
 FT TRANSSEM 81216 81416  
 FT TRANSSEM 81416 81616  
 FT TRANSSEM 81616 81816  
 FT TRANSSEM 81816 82016  
 FT TRANSSEM 82016 82216  
 FT TRANSSEM 82216 82416  
 FT TRANSSEM 82416 82616  
 FT TRANSSEM 82616 82816  
 FT TRANSSEM 82816 83016  
 FT TRANSSEM 83016 83216  
 FT TRANSSEM 83216 83416  
 FT TRANSSEM 83416 83616  
 FT TRANSSEM 83616 83816  
 FT TRANSSEM 83816 84016  
 FT TRANSSEM 84016 84216  
 FT TRANSSEM 84216 84416  
 FT TRANSSEM 84416 84616  
 FT TRANSSEM 84616 84816  
 FT TRANSSEM 84816 85016  
 FT TRANSSEM 85016 85216  
 FT TRANSSEM 85216 85416  
 FT TRANSSEM 85416 85616  
 FT TRANSSEM 85616 85816  
 FT TRANSSEM 85816 86016  
 FT TRANSSEM 86016 86216  
 FT TRANSSEM 86216 86416  
 FT TRANSSEM 86416 86616  
 FT TRANSSEM 86616 86816  
 FT TRANSSEM 86816 87016  
 FT TRANSSEM 87016 87216  
 FT TRANSSEM 87216 87416  
 FT TRANSSEM 87416 87616  
 FT TRANSSEM 87616 87816  
 FT TRANSSEM 87816 88016  
 FT TRANSSEM 88016 88216  
 FT TRANSSEM 88216 88416  
 FT TRANSSEM 88416 88616  
 FT TRANSSEM 88616 88816  
 FT TRANSSEM 88816 89016  
 FT TRANSSEM 89016 89216  
 FT TRANSSEM 89216 89416  
 FT TRANSSEM 89416 89616  
 FT TRANSSEM 89616 89816  
 FT TRANSSEM 89816 90016  
 FT TRANSSEM 90016 90216  
 FT TRANSSEM 90216 90416  
 FT TRANSSEM 90416 90616  
 FT TRANSSEM 90616 90816  
 FT TRANSSEM 90816 91016  
 FT TRANSSEM 91016 91216  
 FT TRANSSEM 91216 91416  
 FT TRANSSEM 91416 91616  
 FT TRANSSEM 91616 91816  
 FT TRANSSEM 91816 92016  
 FT TRANSSEM 92016 92216  
 FT TRANSSEM 92216 92416  
 FT TRANSSEM 92416 92616  
 FT TRANSSEM 92616 92816  
 FT TRANSSEM 92816 93016  
 FT TRANSSEM 93016 93216  
 FT TRANSSEM 93216 93416  
 FT TRANSSEM 93416 93616  
 FT TRANSSEM 93616 93816  
 FT TRANSSEM 93816 94016  
 FT TRANSSEM 94016 94216  
 FT TRANSSEM 94216 94416  
 FT TRANSSEM 94416 94616  
 FT TRANSSEM 94616 94816  
 FT TRANSSEM 94816 95016  
 FT TRANSSEM 95016 95216  
 FT TRANSSEM 95216 95416  
 FT TRANSSEM 95416 95616  
 FT TRANSSEM 95616 95816  
 FT TRANSSEM 95816 96016  
 FT TRANSSEM 96016 96216  
 FT TRANSSEM 96216 96416  
 FT TRANSSEM 96416 96616  
 FT TRANSSEM 96616 96816  
 FT TRANSSEM 96816 97016  
 FT TRANSSEM 97016 97216  
 FT TRANSSEM 97216 97416  
 FT TRANSSEM 97416 97616  
 FT TRANSSEM 97616 97816  
 FT TRANSSEM 97816 98016  
 FT TRANSSEM 98016 98216  
 FT TRANSSEM 98216 98416  
 FT TRANSSEM 98416 98616  
 FT TRANSSEM 98616 98816  
 FT TRANSSEM 98816 99016  
 FT TRANSSEM 99016 99216  
 FT TRANSSEM 99216 99416  
 FT TRANSSEM 99416 99616  
 FT TRANSSEM 99616 99816  
 FT TRANSSEM 99816 100016  
 FT TRANSSEM 100016 100216  
 FT TRANSSEM 100216 100416  
 FT TRANSSEM 100416 100616  
 FT TRANSSEM 100616 100816  
 FT TRANSSEM 100816 101016  
 FT TRANSSEM 101016 101216  
 FT TRANSSEM 101216 101416  
 FT TRANSSEM 101416 101616  
 FT TRANSSEM 101616 101816  
 FT TRANSSEM 101816 102016  
 FT TRANSSEM 102016 102216  
 FT TRANSSEM 102216 102416  
 FT TRANSSEM 102416 102616  
 FT TRANSSEM 102616 102816  
 FT TRANSSEM 102816 103016  
 FT TRANSSEM 103016 103216  
 FT TRANSSEM 103216 103416  
 FT TRANSSEM 103416 103616  
 FT TRANSSEM 103616 103816  
 FT TRANSSEM 103816 104016  
 FT TRANSSEM 104016 104216  
 FT TRANSSEM 104216 104416  
 FT TRANSSEM 104416 104616  
 FT TRANSSEM 104616 104816  
 FT TRANSSEM 104816 105016  
 FT TRANSSEM 105016 105216  
 FT TRANSSEM 105216 105416  
 FT TRANSSEM 105416 105616  
 FT TRANSSEM 105616 105816  
 FT TRANSSEM 105816 106016  
 FT TRANSSEM 106016 106216  
 FT TRANSSEM 106216 106416  
 FT TRANSSEM 106416 106616  
 FT TRANSSEM 106616 106816  
 FT TRANSSEM 106816 107016  
 FT TRANSSEM 107016 107216  
 FT TRANSSEM 107216 107416  
 FT TRANSSEM 107416 107616  
 FT TRANSSEM 107616 107816  
 FT TRANSSEM 107816 108016  
 FT TRANSSEM 108016 108216  
 FT TRANSSEM 108216 108416  
 FT TRANSSEM 108416 108616  
 FT TRANSSEM 108616 108816  
 FT TRANSSEM 108816 109016  
 FT TRANSSEM 109016 109216  
 FT TRANSSEM 109216 109416  
 FT TRANSSEM 109416 109616  
 FT TRANSSEM 109616 109816  
 FT TRANSSEM 109816 110016  
 FT TRANSSEM 110016 110216  
 FT TRANSSEM 110216 110416  
 FT TRANSSEM 110416 110616  
 FT TRANSSEM 110616 110816  
 FT TRANSSEM 110816 111016  
 FT TRANSSEM 111016 111216  
 FT TRANSSEM 111216 111416  
 FT TRANSSEM 111416 111616  
 FT TRANSSEM 111616 111816  
 FT TRANSSEM 111816 112016  
 FT TRANSSEM 112016 112216  
 FT TRANSSEM 112216 112416  
 FT TRANSSEM 112416 112616  
 FT TRANSSEM 112616 112816  
 FT TRANSSEM 112816 113016  
 FT TRANSSEM 113016 113216  
 FT TRANSSEM 113216 113416  
 FT TRANSSEM 113416 113616  
 FT TRANSSEM 113616 113816  
 FT TRANSSEM 113816 114016

```

DR PDB; 10L; 15-MAY-97.
DR PDB; 2LIP; 12-MAR-97.
DR PDB; 3LIP; 16-JUN-97.
DR PDB; 1HOD; 22-AUG-01.
DR PDB; 4LIP; 19-AUG-98.
DR InterPro; IPR000073; A/b hydrolase.
DR InterPro; IPR008262; Lipase AS.
DR InterPro; IPR000379; Ser_eستر.
DR Pfam; PF00561; abhydrolase; 1.
DR PROSITE; PS00120; LIPASE_SER; 1.
DR HydroLase; Lipid degradation; Signal; Calcium; 3D-structure.
FT CHAIN 1 44 LIPASE.
FT ACT_SITE 131 364 RELAY SYSTEM.
FT ACT_SITE 286 286 CHARGE RELAY SYSTEM.
FT ACT_SITE 330 330 CHARGE RELAY SYSTEM.
FT TURND 48 49
FT STRAND 55 58
FT TURND 61 62
FT STRAND 66 67
FT TURND 68 70
FT STRAND 71 72
FT TURND 75 76
FT HELIX 77 83
FT TURND 84 85
FT STRAND 88 90
FT TURND 102 103
FT HELIX 105 120
FT TURND 121 121
FT STRAND 125 130
FT TURND 131 132
FT HELIX 133 143
FT HELIX 145 147
FT STRAND 148 154
FT TURND 158 159
FT HELIX 162 173
FT TURND 172 173
FT TURND 175 176
FT HELIX 178 194
FT TURND 196 197
FT HELIX 204 210
FT TURND 211 221
FT HELIX 213 222
FT TURND 226 227
FT STRAND 228 228
FT TURND 231 232
FT STRAND 236 236
FT STRAND 240 242
FT STRAND 247 255
FT STRAND 258 264
FT TURND 265 266
FT STRAND 267 272
FT TURND 273 274
FT TURND 277 279
FT TURND 281 281
FT HELIX 282 282
FT HELIX 287 300
FT TURND 301 303
FT STRAND 307 307
FT STRAND 312 312
FT HELIX 313 316
FT STRAND 320 326
FT TURND 330 331
FT HELIX 332 334
FT TURND 335 339
FT TURND 343 344
FT HELIX 348 362
FT TURND 363 364
SQ SEQUENCE 364 AA; 37494 MW; E9CD2DBF555658E9 CRC64;
Query Match 9.1%; Score 85.5; DB 1; Length 364;
Best Local Similarity 27.4%; Pred. No. 3.7;
Matches 34; Conservative 20; Mismatches 43; Indels 27; Gaps 6;

```

```

QY 3 IIVHGL-----YMGVWHPLSRHKQGYRQTISSNLAID-----EAI 45
DB 55 IIVHGLSGTDKXAGVLEWYWG----IOEDIQQNGATVYVNLGFSQSDDEPNRGEEL 109
QY 46 FRRLDRL--THASPNAVGHSLGIVIRRYLESAPSCETLSHVAIGSPLOGASIVNK 103
DB 110 LAVYKTVLAAATGATKKNLVGHSGGGL-SSRYVAANVP--DLVASVTTICTPRGSEFADF 166
QY 104 IEOL 107
DB 167 VQDV 170

RESULT 8
ATU1_YEAST STANDARD; PRT; 1216 AA.
ID ATU1_YEAST
AC P38360;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable copper-transporting ATPase (EC 3.6.3.4) (Cu(2+)-ATPase).
GN PCAL OR YBR295W OR YBR2112.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95274324; PubMed=7754711;
RA Rad M.R.; Kirchbach L.; Hollenberg C.;
RT "A putative P-type Cu(2+)-transporting ATPase gene on chromosome II
of Saccharomyces cerevisiae."
RL Yeast 10:1217-1225(1994).
CC -1- FUNCTION: Probably involved in copper transport and in the
regulation of cellular copper level.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + Cu(2+) (in) = ADP + phosphate +
Cu(2+) (out).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the cation transport ATPases family (P-type
ATPases). Subfamily IB.
CC -1- SIMILARITY: Contains 1 HMA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z29332; CA82529.1; -.
DR EMBL; Z36164; CA85260.1; -.
DR PIR; S46177; S46177.
DR GenOnline; 138838; -.
DR SGD; S0000499; PCAL.
DR GO; GO:0016020; C:membrane; IMP.
DR GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. .; IMP.
DR InterPro; IPR006403; ATPase-IB1_Cu.
DR InterPro; IPR006416; ATPase-IB_HV.
DR InterPro; IPR001757; ATPase-IB_E2.
DR InterPro; IPR001756; Cu ATPase.
DR InterPro; IPR008250; E1-E2 ATPase reg.
DR InterPro; IPR006121; HeavyMe transpct.
DR InterPro; IPR005834; Metal bind.
DR InterPro; IPR006191; Metal bind.
DR Pfam; PF00403; HMA; 1.
DR Pfam; PF00702; Hydrolase; 1.
DR PRINTS; PR00119; CATATPAB.
DR PRINTS; PR00943; CUATPASE.
DR TIGRPFMS; TIGR01511; ATPase-IB1_Cu; 1.

```



DR TIGR01525; ATPase-1B hvy; 1.  
 DR TIGR01494; ATPase-P-Type; 2.  
 DR PROSITE; PS00154; ATPase\_E1\_E2; 1.  
 DR PROSITE; PS01047; HMA\_1; PALISE\_NEG.  
 DR PROSITE; PS00846; HMA\_2; 1.  
 KW Hydrolyase; Copper transport; Transmembrane; Phosphorylation;  
 Magnesium; ATP-binding; Metal-binding; Copper.  
 FT DOMAIN 1 556  
 FT TRANSMEM 557 578  
 FT DOMAIN 579 592  
 FT TRANSMEM 593 612  
 FT DOMAIN 613 620  
 FT TRANSMEM 621 641  
 FT DOMAIN 642 659  
 FT TRANSMEM 660 680  
 FT DOMAIN 681 808  
 FT TRANSMEM 809 831  
 FT DOMAIN 832 847  
 FT TRANSMEM 848 865  
 FT DOMAIN 866 1161  
 FT TRANSMEM 1162 1181  
 FT DOMAIN 1182 1190  
 FT TRANSMEM 1191 1209  
 FT DOMAIN 1210 1216  
 FT MOD\_RES 701 701  
 FT METL 421 421  
 FT METL 424 424  
 FT METL 1107 1107  
 FT METL 1111 1111  
 SQ SEQUENCE 1216 AA; 131838 MW; B0BA4D60D75F9EA CRC64;  
 Query Match  
 Best Local Similarity 25.1%; Score 85; DB 1; Length 1216;  
 Matches 43; Conservative 27; Mismatches 73; Indels 26; Gaps 8;  
 QY 15 VNHPLSHRLKGYRTQTSYNSLAIDDAIFRLDRSLTASPNALVHSLG-----GL 69  
 DB 893 VAHNTHSHVFD--KTGLTBKGLTVHETV--RQDR--HNSQSLGLTLEGIKHPVSM 944  
 QY 70 VKRYLESRAAPSCETLSHVVAI-GSPLOGASTVKNIEQLGVALGNSAEGLEKHDV-- 126  
 DB 945 ATASTLKEKGVSAQVNSKATGKRVBETS-----YSLKLGQGNCRMTGHNDPVR 998  
 QY 127 ---ESRYPOKSGSIAGTIPGLRSLILRDPDSDGTVEETKTAGMTDHI 174  
 DB 999 KALGQGVSVFCFSVNGSV---TAVVALBDSLADAVSTINILRQRGISLHI 1046  
 RESULT 9  
 PIRB\_ORYSA STANDARD; PRT; 268 AA.  
 AC Q43360;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE PIR7B protein.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Indica-IR36;  
 RA Reihmann C., Mauch F., Dudler R., Hofmann C.;  
 RT "Characterization of a rice gene induced by Pseudomonas syringae pv.  
 RT Physiol. Mol. Plant Pathol. 46:71-81(1995).  
 CC -1- INDUCTION: By infection of leaves with certain strains of P.  
 CC syringae pv. syringae.  
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B  
 CC CARBOXYL-ESTERASE/LIPASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z34271; CAAB4026.1; -  
 CC EMBL; Z34270; CAAB4024.1; -  
 CC PIR; S47087; S47087.  
 CC HSSP; PS2704; 1074.  
 CC Gramene; Q43360; -  
 CC InterPro; IPR000073; A/b\_hydrolyase.  
 CC InterPro; IPR000379; Ser\_esters.  
 CC Pfam; PF00561; abhydrolyase; 1.  
 KW Hydrolyase; Serine esterase.  
 FT ACT\_SITE 86 86  
 FT ACT\_SITE 218 218  
 FT ACT\_SITE 246 246  
 SQ SEQUENCE 268 AA; 28835 MW; 9F4E67E1216560A1 CRC64;  
 Query Match  
 Best Local Similarity 8.6%; Score 80.5; DB 1; Length 268;  
 Matches 47; Conservative 20; Mismatches 86; Indels 33; Gaps 6;  
 QY 4 ILVHGLYMGVLVNHPLSHRLKGYRTQTSY-----NSLAIDDAIFRLDRSLTHASF 58  
 DB 12 ILVHGLCHGAKCVRVYALRAAGHATYALDMAASGAMHARVDVETFEYSPLDDAYA 71  
 QY 59 NA-----LVHSLDGLVIRKYLESRAPSCETLSHVVAIGSPLOGASTVKNIEQL-- 107  
 DB 72 AAAPGRVLVGHSHGSLVALAME-RPPDKVAALVFAAAMPVCGKMGVPTSEFMER 130  
 QY 108 ----GL-----GVALGNSAEPGLKENDDSRYQKSGSIAGTIPGLRSLILR----- 151  
 DB 131 TAPEGLMDCEMVAINNQSQGVAINLGPFLAQKTYQOSPAEDDLAKYLVAPNGQFMD 190  
 QY 152 DPLDSD 157  
 DB 191 DPLDSD 196  
 RESULT 10  
 PMPC\_CHLMU STANDARD; PRT; 1460 AA.  
 AC O9PUY1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable outer membrane protein pmc precursor (Polymorphic membrane  
 DE protein C).  
 GN PMPC OR TC0695.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OC NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mopn / N199;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Baas S.,  
 RA Linher K., Weidman J., Kouri H., Craven B., Bowman C., Dodson R.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39.";  
 RT Nucleic Acids Res. 28:1397-1406(2000).  
 CC -1- SUBCELLULAR LOCATION: Cell wall surface (elementary bodies)  
 CC (potential).  
 CC -1- SIMILARITY: Belongs to the PMP outer membrane protein family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: A5002338; AAF39511.1; -.  
DR PIR: D81675; D81675.  
DR TIGR: TC0695; -.  
DR InterPro: IPR006315; Autotransport.  
DR InterPro: IPR003368; Chlamydia\_PMP.  
DR Pfam: PF02415; Chlamydia\_PMP; 2.  
DR TIGRPFAM: TIGR01414; autotrans\_bar1; 1.  
DR TIGRPFAM: TIGR01376; POMP repeat; 8.  
KM Outer membrane signal, Multigene family, Complete proteome.  
FT SIGNAL 1 20  
FT CHAIN 21 1460 PROBABLE OUTER MEMBRANE PROTEIN PMPC.  
SQ SEQUENCE 1460 AA; 154781 MW; C1F033E7907AC3AC CRC64;

Query Match 8.6%; Score 80.5; DB 1; Length 1460;  
Best Local Similarity 26.1%; Pred. No. 48;  
Matches 47; Conservative 25; Mismatches 69; Indels 39; Gaps 9;

QY 30 TQTSYNSLAIDDEAIFRRLDRSLTHAS-----PNAIVGHSIGLVIKYLESPAPSCETL 85  
DB 490 TSPITQDSASHVAIFRRIASSSQNSENIPNADGSTAGDGSSQSTGSDSSI 549  
QY 86 SRVVAIGSPQASIVNKEIQ-GLGVALGNSAEFGKHEHDESRYPKSGSIAGTI--P 142  
DB 550 NHYIG-GGAIVGEAV--KIENLSGYGTFSNNNA-----QISGSTDIVGGA 595  
QY 143 LGRSLILADPLDSDQTVVEET-----KIAG-----MTDHIISTSYENA 184  
DB 596 IYAKTSLTIDSGNSGCTIFSENTTSKSTTGQVAGGAIFFRSVITITPVPFSSKNSALNA 655

RESULT 11  
ID ARGC\_SULTO STANDARD; PRT; 349 AA.  
AC 097655;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-  
DE acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).  
GN ARGC OR ST0195.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX MEDLINE=21456156; PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankal A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamida M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagishi M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
RL DNA Res. 8:123-140(2001).  
CC -1- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+)  
CC + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.  
CC -1- PATHWAY: Arginine biosynthesis; third step.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -1- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily  
CC 1.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: APO0981; BAB65152.1; ALT\_INIT.  
DR HAMAP: MF\_00150; -1.  
DR InterPro: IPR00706; AGPR\_act\_site.  
DR InterPro: IPR000534; Semialdh\_dh.  
DR Pfam: PF01118; Semialdehyde\_dh; 1.  
DR Pfam: PF02774; Semialdehyde\_dhc; 1.  
DR ProDom: PD003765; AGPR\_act\_site; 1.  
DR PROSITE: PS01224; AGCT, PALSE NEG.  
KM Arginine biosynthesis; Oxidoreductase; NADP, Complete proteome.  
FT ACT\_SITE 150 150  
FT ACT\_SITE 150 150  
SQ SEQUENCE 349 AA; 38699 MW; EFLE306BBA5847B CRC64;

Query Match 8.5%; Score 79.5; DB 1; Length 349;  
Best Local Similarity 24.8%; Pred. No. 12;  
Matches 38; Conservative 29; Mismatches 53; Indels 33; Gaps 9;

QY 19 LSHRLHKLGRITQTSYNSLAIDDEAIFRRLDRSL-----THASPNALVGHSLGGLVI 71  
DB 76 LPHKV-SLEVVPKILEMGIVQIDLSADFRUKDPTLYKIYGEHPYD-LTKKAVYGLPE 133  
QY 72 KRYLESR-----APSCETLSHVAIGSPQASIVNKEIQGLGVALGNSAEFGKHEH 125  
DB 134 LHYELKNAKLTPSCGNATATTLA-GAPLVASLLETKLIS-DYKVG-SSSGAKKPH 190  
QY 126 DESRYPQKSGSINGTTPGLRSLILADPLDSDQ 158  
DB 191 G-SHHPERQNAIR-----PYEADG 208

RESULT 12  
ID OM40\_HUMAN STANDARD; PRT; 361 AA.  
AC 096008; Q86VW4; Q8WY09; Q8WY10; Q8WY11; Q9BR95;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable mitochondrial import receptor subunit TOM40 homolog  
DE (protein) (p38.5).  
DE (translocase of outer membrane 40 kDa subunit homolog) (Haymaker  
DE protein) (p38.5).  
GN TOM40 OR TOM40 OR PEREC1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=99449047; PubMed=10520737;  
RA Freitas E.M., Zhang W.J., Lalonde J.P., Tay G.K., Gaudieri S.,  
RA Ashworth L.K., Van Bockmeier F.M., Dawkins R.L.;  
RT "Sequencing of 42kb of the APO E-C2 gene cluster reveals a new gene:  
RT PEREC1.";  
RL DNA Seq. 9:89-101(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Yoshizawa K., Murray J.C.;  
RT "A transcriptional map in the region of 19q13 derived using direct  
RT sequencing and exon trapping.";  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Lymphocytes;  
RX MEDLINE=21610875; PubMed=11745481;  
RA Das B., Tao S.-Z., Mubhutsky R., Norin A.J.;  
RT "Genetic identity and differential expression of p38.5 (Haymaker) in  
RT human malignant and non-malignant cells.";  
RL Int. J. Cancer 94:800-806(2001).

[4] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RA TISSUE=EYE, lung, Skin, Testis, and Uterus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stopleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson W.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Schermer A., Schein J.E., Jones S.J.M., Maria M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Essential for the import of protein precursors into the  
 CC mitochondria (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC outer membrane (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=O96008-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O96008-2; Sequence=VSP\_008589, VSP\_008590;  
 CC Note=No experimental confirmation available;  
 CC -1- SIMILARITY: Belongs to the Tom40 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

FT CONFLICT 49 49 S -> R (IN REF. 4; AA06413).  
 FT CONFLICT 219 219 A -> S (IN REF. 3; AA146625).  
 FT CONFLICT 220 220 H -> R (IN REF. 3; AA146624).  
 FT CONFLICT 258 258 N -> S (IN REF. 3; AA146626).  
 FT CONFLICT 297 297 T -> A (IN REF. 3; AA146624).  
 FT CONFLICT 313 313 L -> F (IN REF. 3; AA146624).  
 SQ SEQUENCE 361 AA; 37893 MW; CFE55E01F800D32 CRC64;  
 Query Match 8.5%; Score 79.5; DB 1; Length 361;  
 Best Local Similarity 23.2%; Pred. No. 12;  
 Matches 43; Conservative 32; Mismatches 65; Indels 45; Gaps 8;  
 QY 11 MGLVWEPYSLHRL-KLGRYOTISYNSLAIDPAFLRDLSTPLASPNALVGHSLGL 69  
 DB LAAQVTHQGPLRBRKMAIQTSKFWNQDGEYRSGDFPLAATLGNPDVVG---SGI 216  
 QY 70 VKRYTES-----RAPSCE-----TLSH---VVAIGSLQASIVN 102  
 DB 217 LVAAHYLGSTPCLALGSLVYHRRRGEQTVWSLAGKYTLNNWTLATVTLGQGMATVYH 276  
 QY 103 KI-EOLGVLGALNAGAEKGVKHEHDESRYPQSGSIAGTIPILGRLILRDLSDGY- 160  
 DB 277 KASDQQLQGV-----EFASITMQDTSVFGYQLDLPKRNLLFKGSVDSNWIVG 325  
 QY 161 -TVBR 164  
 DB 326 ATLEK 330  
 RESULT 13  
 ACVS EMENT  
 ID ACVS EMENT STANDARD; PRT; 3770 AA.  
 AC P27742;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 43, Last annotation update)  
 DE N-(5-amino-5-carboxypentanoyl)-L-cysteiny-D-valine synthase  
 DE (EC 6.3.2.26) (Delta-(L-alpha-aminoacyl)-L-cysteiny-D-valine  
 DE synthetase) (ACV synthetase) (ACVS).  
 GN ACVA.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxId=162425;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=9191;  
 RX MEDLINE=91286299; PubMed=2061333;  
 RA Maccabe A.P., van Lierp H., Pallissa H., Unkles S.E., Riach M.B.R.,  
 RA Pfeiffer E., von Doehren H., Kinghorn J.R.,  
 RA Aspergillus nidulans. Molecular characterization of the acvA gene  
 RT encoding the first enzyme of the penicillin biosynthetic pathway".  
 RL J. Biol. Chem. 266:12646-12654(1991).  
 CC -1- FUNCTION: Each of the constituent amino acids of the tripeptide  
 CC acv are activated as aminoacyl-adenylates with phosphate bonds  
 CC formed through the participation of amino acid thioester  
 CC intermediates.  
 CC -1- CATALYTIC ACTIVITY: L-2-aminohexanoate + L-cysteine + L-valine  
 CC + 3 ATP = N-(L-5-amino-5-carboxypentanoyl)-L-cysteiny-D-valine +  
 CC 3 AMP + 3 diphosphate.  
 CC -1- COFACTOR: Contains 3 covalently bound phosphoguanethamines  
 CC (Potential).  
 CC -1- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.  
 CC -1- PTM: The N-terminus is blocked.  
 CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme  
 CC family.  
 CC -1- SIMILARITY: Contains 3 acyl carrier domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X54853; CAA38631.1; -  
DR PIR; A40889; A40889.  
DR HSP; P14687; 1AMU.  
DR InterPro; IPR000873; AMP-bind.  
DR InterPro; IPR001242; Condensatn.  
DR InterPro; IPR006163; Pp\_bind.  
DR InterPro; IPR006162; Pentose\_S.  
DR InterPro; IPR000379; Ser\_estrs.  
DR InterPro; IPR001031; Thioesterase.  
DR Pfam; PF00501; AMP-binding; 3.  
DR Pfam; PF00668; Condensation; 3.  
DR Pfam; PF00550; pp-binding; 3.  
DR Pfam; PF00975; Thioesterase; 1.  
DR PRINTS; PR00154; AMPBINDING.  
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.  
DR PROSITE; PS00455; AMP\_BINDING; 3.  
DR PROSITE; PS00075; ACP\_DOMAIN; 3.  
KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;  
Repeat; Phosphopantetheine.  
KW Repeat; Phosphopantetheine.  
FT REPEAT 321 910 DOMAIN 1 (ADIPATE-ACTIVATING).  
FT REPEAT 1413 1993 DOMAIN 2 (CYSTEINE-ACTIVATING).  
FT REPEAT 2494 3078 DOMAIN 3 (VALINE-ACTIVATING).  
FT DOMAIN 850 919 ACYL CARRIER (ACP) 1.  
FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.  
FT DOMAIN 3087 3087 ACYL CARRIER (ACP) 3.  
FT BINDING 882 882 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
FT BINDING 1965 1965 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
FT BINDING 3050 3050 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
FT ACT SITE 3623 3623 THIOESTERASE (BY SIMILARITY).  
SQ SEQUENCE 3770 AA; 422448 MW; C866B6D232A58CB0 CRC64;

Query Match 8.5%; Score 79.5; DB 1; Length 3770;  
Best Local Similarity 25.3%; Pred. No. 1.7e+02; Indels 31; Gaps 8;  
Matches 42; Conservative 31; Mismatches 62; Indels 31; Gaps 8;

QY 19 LSHRLKLYRTQTSYNSLAIDDAIFR-RIDRSLTASPNALVGHSLGLVIRKYLE 77  
DB 1971 LSFKHVEVGRITISV-----ALFRHRTISLAI-----LIMNVGD-----IQE 2010  
QY 78 RAPSETSHVVAIGSPQASIVNKIEQLGVALGSAFGLKEHDERYPOKSGSI 137  
DB 2011 ITPVDYDNRKTAIVSPAQERLLFIHELH--GGGNAVNIIDAFELPPYIDQSRVEALYTI 2068  
QY 138 ACTIPGLRSLTRDPLDSDGT-----VTVEETKLAGMTDIAIST 178  
DB 2069 LSRHE-ALRTFLIRD--QATGTFYOKITITPDAKMLIKSNVST 2111

RESULT 14  
YCSB\_BACSU STANDARD; PRT; 249 AA.  
AC P42962;  
DT 01-NOV-1995 (rel. 32, Created)  
DT 01-NOV-1995 (rel. 32, Last sequence update)  
DT 10-OCT-2003 (rel. 42, Last annotation update)  
DE Hypothetical protein ycsb.  
OS YCSB OR BSU04040.  
GN Bacillus subtilis.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_Taxid:1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=168;  
RA Akagawa E., Kurita K., Sugawara T., Nakamura K., Yamane K.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RC SEQUENCE FROM N.A.  
RC STRAIN=168;  
RC MEDLINE=97124189; PubMed=8969502;

RA Yamane K., Kumano M., Kurita K.;  
RT "The 25 degrees-36 degrees region of the Bacillus subtilis  
RT chromosome: determination of the sequence of a 146 kb segment and  
RT identification of 113 genes.";  
RL Microbiology 142:3047-3056 (1996).

RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RC MEDLINE=98044033; PubMed=9384377;  
EX Kunst F., Ogawara N., Moser I., Albertini A.M., Alloni G.,  
RA Arevalo V., Bortier M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Bortier L., Bortier L., Brans A., Braun M., Brignall S.C., Bron S.,  
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.D., Comercon I.F., Cummings N.J., Daniel R.A.,  
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Ertan K.D., Errington J., Fabret C., Ferrati E., Fougere D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Gim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,  
RA Gilsenpl G., Guy B.J., Haga K., Halech U., Harwood C.R., Henaut A.,  
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Ilaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaer-Bianhard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapierre A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Meliadi R.P., Mizuno M., Noesti D., Nakai S., Nodack M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,  
RA Paro V., Pohl T.M., Portetlelle D., Portolillo S., Prescott A.M.,  
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scofield S.,  
RA Sekiguchi J., Sekowaka A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsita P., Tognoni A.,  
RA Toseco V., Uchiyama S., Vandembol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Welzenegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein E.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256 (1997).  
CC - SIMILARITY: BELONGS TO THE COF/YBHA/YIDA/YTGL (E.COLI) / YCSB/YXEH  
CC (B.SUBTILIS) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D38161; BAA07356.1; -  
CC EMBL; D50453; BAA09035.1; -  
CC EMBL; Z99106; CAB12212.1; -  
CC PIR; I39893; I39893.  
CC Subtilisin; BG11226; YCSB.  
CC InterPro; IPR001757; ATPase\_E1-E2.  
CC InterPro; IPR006379; HAD\_SF\_IIB.  
CC InterPro; IPR005834; Hydrolase.  
CC InterPro; IPR000150; Hypoet\_cof.  
CC Pfam; PF00702; Hydrolase; 1.  
CC PRINTS; PR00119; CATALPASE.  
CC TIGRFAMs; TIGR00099; Cof-subfamily; 1.  
CC TIGRFAMs; TIGR01484; HAD-SF-IIB; 1.  
CC PROSITE; PS01228; COF\_1; 1.  
CC PROSITE; PS01229; COF\_2; 1.  
CC Hypothetical protein; Complete proteome.  
KW SEQUENCE 249 AA; 28068 MW; C6D420A8EC08C57 CRC64;

Query Match 8.4%; Score 79; DB 1; Length 249;  
Best Local Similarity 26.4%; Pred. No. 8.7;  
Matches 24; Conservative 22; Mismatches 27; Indels 18; Gaps 4;

```

QY 43 EAIFRRLRLSLTHASPNALVGHSLGLVTKRYKLESPASCSCE-----TLSHVVALGSPSL 95
DB 150 BELRTKTSSEITNSPNTNIEVNALG-----INKAAPCGYCKLGTFTWENVAMWDSL 202
QY 96 QGASIVNKIEQLGLGVALGNSAERGLKEHDD 126
DB 203 NDIMW---IKKAGLGVMAGNAGDI-VKETAD 229

RESULT 15
1963_MYCTU STANDARD; PRT; 266 AA.
AC P71547;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein RV0963c/MT0992/MD0988c.
GN RV0963c OR MT0992 OR MTCY10D7.11 OR MB0988c.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OK NCBI_taxid=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.B. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean L.A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.B., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218035;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kojanay J.F., Nelson W.C., Mayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12786972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -1- SIMILARITY: SOME. TO M.TUBERCULOSIS RV2079.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; Z79700; CAB02008.1; -.
CC DR EMBL; AE006963; AKA45240.1; -.
CC DR EMBL; BX248337; CAD93849.1; -.

```

```

DR PIR; A70718; A70718.
DR TIGR; MT0992; -.
KW Tuberculosis; RV0963c; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 266 AA; 28069 MW; E1FA43081A6B61AD CRC64;

Query Match      8.4%; Score 79; DB 1; Length 266;
Best Local Similarity 26.2%; Pred. No. 9.3;
Matches 42; Conservative 19; Mismatches 51; Indels 40; Gaps 9;

QY 26 LGYRT-----QTISYNLAIDRAIFRRLDRSL---THASPNALV---GSLGLGVTKRYLE 76
DB 122 LGYDAPDGLKDYVHDMWSARDAAGPLNRPDKGLAATTNVSDDCHITAFGHSYGSLSLALQ 181
QY 77 SRAPSCETLSHYVALGSPVQASIVNKIEQLGLGVALGNS-AERGLKEHDDSRYPQKSG 135
DB 182 QGAP---VSDVVLVGGSP---GTSLTHASQ---LGVEPGHAFYMGVNDH----- 221
QY 136 SIAGTIP-----LGLRSLDRDPLDSPG 158
DB 222 -VANTIFERGARGSAPQDVPGMTQLSVNTGLAPGFLSDG 260

```

Search completed: April 29, 2004, 11:36:01  
Job time : 35 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 11:24:52 ; Search time 41 Seconds  
(without alignments)  
1423.680 Million cell updates/sec

Title: US-10-603-260-5  
Profile score: 936  
Sequence: 1 MGIILVHGLVHGLVHPLS.....KIAQMDHIAISTSYENAV 185

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_rhiz:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	577.5	61.7	215	16	Q8DSU8
2	566.5	60.5	205	16	Q87GR6
3	532	56.8	209	16	Q9KH73
4	179.5	19.2	199	16	Q8PEE2
5	170.5	18.2	224	16	Q89S87
6	155	16.6	98	2	Q87026
7	136.5	14.6	328	16	Q7UNU3
8	136	14.5	180	16	Q8PHU7
9	121.5	13.0	249	16	Q55748
10	112.5	12.0	251	16	Q9PHH3
11	111.5	11.9	254	2	Q8KU33
12	107	11.4	256	16	Q98KTS
13	105.5	11.3	233	16	Q98UB5
14	105.5	11.3	288	16	Q8UXB1
15	104.5	11.2	243	16	Q92XY1
16	103.5	11.1	202	16	P73372

17	102.5	11.0	256	16	Q89ML2	Q89ml2 bradyrhizob
18	99.5	10.6	211	16	Q8YK63	Q8YK63 anabaena sp
19	97	10.4	293	2	Q9SV86	Q9SV86 pseudomonas
20	97	10.4	309	16	Q87QG8	Q87QG8 vibrio para
21	97	10.4	623	16	Q8F4V3	Q8F4V3 leptospira
22	96.5	10.3	309	2	P72172	P72172 pseudomonas
23	96.5	10.3	309	16	Q9HUZ7	Q9HUZ7 pseudomonas
24	96	10.3	327	16	Q9ZBN4	Q9ZBN4 streptomyces
25	95.5	10.2	340	16	Q8X873	Q8X873 escherichia
26	95.5	10.2	340	16	Q8FCX8	Q8FCX8 escherichia
27	95.5	10.2	340	16	Q83JB6	Q83JB6 shigella fl
28	95	10.1	255	16	Q8UKN4	Q8UKN4 agrobacteri
29	95	10.1	262	16	Q7UOF1	Q7UOF1 rhodospirill
30	94.5	10.1	259	16	Q8XWZ0	Q8XWZ0 ralsconia s
31	94.5	10.1	261	16	Q8XWZ0	Q8XWZ0 bradyrhizob
32	94.5	10.1	338	2	Q7X568	Q7X568 acinetobact
33	94.5	10.1	340	16	Q7UAR6	Q7UAR6 shigella fl
34	94	10.0	332	16	Q916C6	Q916C6 pseudomonas
35	93.5	10.0	336	13	Q7ZTW6	Q7ZTW6 xenopus lae
36	93	9.9	302	16	Q8DA61	Q8DA61 vibrio vuln
37	92.5	9.9	258	16	Q81EC2	Q81EC2 bacillus ce
38	92.5	9.9	284	16	Q8KQJ2	Q8KQJ2 vibrio chol
39	92	9.8	290	16	Q8S295	Q8S295 streptomyces
40	91.5	9.8	344	13	Q7E314	Q7E314 brachydanio
41	91	9.7	242	10	Q9FF27	Q9FF27 arabidopsis
42	91	9.7	266	16	Q89R93	Q89R93 bradyrhizob
43	90.5	9.7	265	16	Q82HP7	Q82HP7 streptomyces
44	90.5	9.7	300	16	Q92D71	Q92D71 listeria in
45	90	9.6	195	16	Q7U656	Q7U656 synechococc

#### ALIGNMENTS

RESULT 1	Q8DSU8	PRELIMINARY;	PRT;	215 AA.
ID	Q8DSU8			
AC	Q8DSU8			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Predicted acetyltransferase/hydrolase.			
GN	V20805.			
OS	Vibrio vulnificus.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;			
OC	Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=672;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CMCP6;			
RA	Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,			
RA	Choy H.E.;			
RT	"Complete genome sequence of Vibrio vulnificus CMCP6.";			
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AB016810; AAC07731.1; -			
DR	GO: GO:0016787; F:hydrolase activity; IEA.			
DR	GO: GO:0016740; F:transferase activity; IEA.			
DR	InterPro: IPR000379; Ser. esters.			
KW	Hydrolase; Transferase; Complete proteome.			
SQ	SEQUENCE 215 AA; 23875 MW; 5AD7E9FE1B690E2E CRC64;			

Query Match	61.7%;	Score 577.5;	DB 16;	Length 215;
Best Local Similarity	64.6%;	Pred. No. 1.2e-45;		
Matches 113;	Conservative 28;	Mismatches 33;	Indels 1;	Gaps 1;
QY	1	MGIILVHGLVHGLVHPLSHRLKGYRTQTSYNSLAIDDEALFRRLDRSLTASPNV	60	
DB	6	MGIILVHGLVHGLVHPLSHRLKGYRTQTSYNSLAIDDEALFRRLDRSLTASPNV	65	
QY	61	LVGHSLGIMIRHYLASRPTLQQLSHVVALGSPFKGASIVTKDELGLGAMLGNSPKHG	125	
DB	66	LVGHSLGIMIRHYLASRPTLQQLSHVVALGSPFKGASIVTKDELGLGAMLGNSPKHG	125	

QY 121 LKEHDESRYPKSGSIAGTIPLGRLSLRLDPLD-SPGTVVEETKIAGMTDHI 174  
 DB 126 LAMHEDVWDFPKLGSIGAGTIPLGARPLPLMDTQMSDGVIVETIKSGMTDHY 180

RESULT 2  
 ID 087GR6 PRELIMINARY; PRT; 205 AA.  
 AC 087GR6;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-OCT-2003 (TRENBLrel. 24, Last sequence update)  
 DE Hypochemical protein.  
 GN VPAL249.  
 OS *Vibrio parahaemolyticus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrrio.  
 NC NCB1\_TaxID=670;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RMD 2210633 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Najiima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shinagawa H., Hattoni M., Iida T.,  
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism  
 RT distinct from that of *V. cholerae*.";  
 RL Lancet 361:743-749 (2003).  
 DR EMBL, AF005088; F6C62592.1;  
 DR GO; GO:0003824; F6C62592.1; Ser. esters.  
 DR InterPro; IPR000379; Ser. esters.  
 KW Hypochemical protein; Complete proteome.  
 SQ SEQUENCE 205 AA; 22594 MW; F6A9AE2DC67A2539 CRC64;

Query Match 60.5%; Score 566.5; DB 16; Length 205;  
 Best Local Similarity 61.1%; Pred. No. 1.1e-44;  
 Matches 110; Conservative 27; Mismatches 42; Indels 1; Gaps 1;

QY 1 MOIIVHGLVHGLVHPLSHRLHKLGYRTQITSYNSLAIDDEALFRRLDRLSLTHASPA 60  
 DB 1 MKIIVHGLVHGLVHPLSHRLHKLGYRTQITSYNSLAIDDEALFRRLDRLSLTHASPA 60

QY 61 LVGHSIGLVIKRYESRAPSCETLSHYVAISPIQGASIVNKIEQLGVALGNSAEFG 120  
 DB 61 LVGHSIGLVIKRYESRAPSCETLSHYVAISPIQGASIVNKIEQLGVALGNSAEFG 120

QY 121 LKEHDESRYPKSGSIAGTIPLGRLSLRLDPLD-SPGTVVEETKIAGMTDHI 174  
 DB 121 LKHHDMAFPKLGSIGAGTIPLGARPLPLMDTQMSDGVIVETIKSGMTDHY 180

RESULT 3  
 ID 09KN73 PRELIMINARY; PRT; 209 AA.  
 AC 09KN73;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DE Hypochemical protein VCA0092.  
 GN VCA0092.  
 OS *Vibrio cholerae*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrrio.  
 NC NCB1\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Brimacombe M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
 RT *cholerae*.";  
 RL Nature 406:477-483 (2000).  
 DR EMBL; AE004352; AAF96006.1;  
 DR PIR; B82501; B82501.  
 DR TIGR; VCA0092;  
 DR GO; GO:0003824; F6C62592.1; Ser. esters.  
 DR InterPro; IPR000379; Ser. esters.  
 KW Hypochemical protein; Complete proteome.  
 SQ SEQUENCE 209 AA; 23154 MW; 4747B6B8E54DA08E CRC64;

Query Match 56.8%; Score 532; DB 16; Length 209;  
 Best Local Similarity 59.1%; Pred. No. 1.9e-41;  
 Matches 104; Conservative 26; Mismatches 44; Indels 0; Gaps 0;

QY 1 MOIIVHGLVHGLVHPLSHRLHKLGYRTQITSYNSLAIDDEALFRRLDRLSLTHASPA 60  
 DB 1 MKIIVHGLVHGLVHPLSHRLHKLGYRTQITSYNSLAIDDEALFRRLDRLSLTHASPA 60

QY 61 LVGHSIGLVIKRYESRAPSCETLSHYVAISPIQGASIVNKIEQLGVALGNSAEFG 120  
 DB 61 LVGHSIGLVIKRYESRAPSCETLSHYVAISPIQGASIVNKIEQLGVALGNSAEFG 120

QY 121 LKEHDESRYPKSGSIAGTIPLGRLSLRLDPLD-SPGTVVEETKIAGMTDHI 174  
 DB 121 LKHQDSWELPGLGSIAGTIPLGARPLPLMDTQMSDGVIVETIKSGMTDHY 180

RESULT 4  
 ID 08P6E2 PRELIMINARY; PRT; 199 AA.  
 AC 08P6E2;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DE Hypochemical protein XCG3028.  
 GN XCG3028.  
 OS *Xanthomonas campestris* (pv. *campestris*).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 NC NCB1\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPPB 528;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Chagoff R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Camarvan F., Cardozo J., Chambergo F., Chapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
 RA Faria J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Serubal J.C., Kiteajima J.P.,  
 RT "Comparison of the genomes of two *Xanthomonas* pathogens with differing  
 RT host specificities.";  
 RL Nature 417:459-463 (2002).  
 DR EMBL; AB012415; AA042289.1;  
 KW Hypochemical protein; Complete proteome.  
 SQ SEQUENCE 199 AA; 20895 MW; 2E671047A861570 CRC64;

Query Match 19.2%; Score 179.5; DB 16; Length 199;  
 Best Local Similarity 31.8%; Pred. No. 1.2e-08;  
 Matches 57; Conservative 29; Mismatches 86; Indels 7; Gaps 4;



```

QY 3 IIVHAGIYMGGLWHPFSHLLHGLRTRCTISNLSAIDEALFRDLSDSLTHASPNLV 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MLVTHGIMTWHTMLPLFARLRTRAGMTPLFGVAVRGSGPAQVPLIRLRATRAAPLV 60

QY 63 GHSIGIGVIRKYLESPACETLSHVVAAGSPLOQASINNKLEOGLGVALNSAPFLX 12
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 CHSIGGIMALQNG-QAPFL-PYQRVVCLOSPICSSAARGLRAGRGHMAQRSALILIR 11

QY 123 --EHDDSRKFPQSGSAGTIPICGLBSLIRDLBDDGYVWYBETKIAMTDHIAISTT 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 GFTMDA---AQQGQAGCVRPGRLGRWLAFLQAGSDGVYAGDETRGLPELHHCVQAS 174

```

## RESULT 5

ID	Q89SS7	PRELIMINARY;	PRT;	224	AA
----	--------	--------------	------	-----	----

PT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE B112323 protein.  
GN B112323.  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_taxid=575;  
RX [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=USDA\_110;  
RC MEDLINE=22484998; PubMed=12597275;  
RX Kaneo T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Ideesawa K., Iiguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada  
RA Tabata S.;  
RT "complete genomic sequence of nitrogen-fixing symbiotic bacterium  
RT Bradyrhizobium japonicum USDA110.";  
RL DNA Res. 9:189-197 (2002).  
EMBL; AP005843; BAC47588.1; "-  
DR GO; GO:0003824, Fcatalytic activity, IEA.  
DR InterPro; IPR000379; Set\_restrict.  
KW Complete proteome.  
SQ SEQUENCE 224 AA; 23746 MW; D58F7229ED11465 CRC64;

Query Match	18.2%;	Score 170.5;	DB 16;	Length 224;
Best Local Similarity	29.1%;	Pred. No. 9.5e-08;		
Matches 57; Conservative	34;	Mismatches 74;	Indels 31;	Gaps 8;

```

QY      3      IIIIVGGIYMGGLVWHEPLSHRLKLGRTGTTISYNS-----LAIDSE--ALFRRLDSLSLT 54
DQ      15      VVLHGIRARSASVLTLELRALRASGTTLVNDVYSRSRKPALADDDHPALARAEND-- 72
QY      55      HAAEPNNLVGHSILGGLIVIKRYLESRAESCETSHVAALSPLOGASIVNKIEQL----- 107
DQ      73      --APLPHVAHSMGGLVARAYIARHRA--RLARVVLGTPNNGSEVADLLGGLLPYPAFY 128
QY      108     -GGVVALGN---SAEFGLEKHEDDESKYPOKSSISGTTPLGLRSLLLRPDPSDGYTVE 163
DQ      129     GPAGLELTETTPPAAPNALPAID---YP--VGVIAGNRFIDPVAGLFVLPPNDGSRVQ 182
QY      164     ETKIAGMTDIIAISTT 179
DQ      183     STVLAGMTDIIHVVKAS 198

```

## RESULT 6

ID	087026	PRELIMINARY;	PRT;	98 AA
----	--------	--------------	------	-------

DT 01-NOV-1998 (TRENDSrel. 08, Created)  
DT 01-NOV-1998 (TRENDSrel. 08, Last sequence update)  
DT 01-NOV-1998 (TRENDSrel. 08, Last annotation update)  
DE Z10r protein.  
GN Z10r.

OS *Vibrio cholerae*.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales  
OC Vibrionaceae; Vibrrio.  
OX NCBI\_TaxID=666;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=217561;  
RC Fallerino A.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ231080; CAAL312.1; -  
QX SEQUENCE 98 AA; 11040 NM; 4350024076808FP9C CRC64;

Query Match	16.6%;	Score 155;	DB 2;	Length 98;
Best Local Similarity	50.0%;	Pred. No. 8.5e-07;		

<b>Qy</b>	113 LGNSAFPLGKHEHDESRYPKSGSIAGRIPGLRSLILRPDLPSDGVVVEETKIAGMTD	172
	:::::     :::     :::     :::     :::     :::     :::     :::	
<b>Db</b>	2 LGSNAHYTGLQHDSMELPQRIGCTIAGRIRGRFPRLDGSSGMCDetIVVAERQISGMTD	61
<b>Qy</b>	173 HIAI 176	
	:::	
<b>Db</b>	62 HLIL 65	

## RESULT 7

ID	Q7JUN03	PRELIMINARY;	PRT;	328	AA
----	---------	--------------	------	-----	----

DT 01-OCT-2003 (T<sub>EMBL</sub>rel. 25, Created)  
DT 01-OCT-2003 (T<sub>EMBL</sub>rel. 25, Last sequence update)  
DT 01-OCT-2003 (T<sub>EMBL</sub>rel. 25, Last annotation update)  
DE Hypothetical protein.  
GN RB374.  
OS Rhodospirillum rubrum.  
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
OC Planctomycetaceae; Pirellula.  
OX NCBI\_TaxId=117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=1;  
RC  
RX MEDLINE=82735913; PubMed=12835416;  
RA Gloeckner F.O., Kobe M., Bauer M., Teeling H., Lombardot T.,  
Rudolf W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,  
RA Schlessner H., Aumann R., Reinhardt R.;  
RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
RT strain 1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:8238-8303 (2003).  
DR EMBL: BX294146; CAD75322.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 328 AA; 3608 MW; 6FB831D70FCFEF59 CRC64;

Query Match	14.6%;	Score 136.5;	DB 16;	Length 328;
Best Local Similarity	27.8%;	Pred. No. 0.00024;		
Matches 52;	Conservative 34;	Mismatches 74;	Indels 27;	Gaps 9

```

QY      3  IILVHVLNHHGVMHPLSLRLKIGYRROTTSY-----NSLAIDDEKIFRLRLRSTHA 56
Dd      118 VVILHGEKMTDNCKMSLETKYLAEGYE-QTIRFGYASTRSDLA-BSAALALADVLENGSSD 172
QY      57  SPNALVGHSLGGVVKRYLYBSAPSCSCE-----TLSHVAALGSPLOGASIVNKIEQLGL- 103
Dd      176 AEPAPFVGHSMGNIVTRPHLIGLDQADGDCKNILLPRKSMVMGLGPPNQGAAIARRLAPGVF 235
QY      110 GVALGNSA-EFGIKHEHDSRY---PQSGSAGIITPLGLASLLRPDL---DDGTYTV 162
Dd      236 GLVAGGGAAMELGRSMVNSHLPATPPFPFAIVAGRY-----ATPLANFLVDGDDDFVSVL 295
QY      163 EETKIG 169
Dd      291 EEAOLEG 297

```

```

RESULT 8
ID 08PHU7 PRELIMINARY; PRT; 180 AA.
AC 08PHU7;
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DE 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Hypothetical protein XAC3152.
GN XAC3152.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Camavari F., Cardoso J., Chamberg F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Curcio-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Gruber A.,
RA Fontignier E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
RA Setudal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011959; AAM37996.1; -.
DR Hypothetical protein, Complete proteome.
KM SEQUENCE 180 AA; 19018 MM; 6DDB43872694A746 CRC64;
SQ
Query Match 14.5%; Score 136; DB 16; Length 180;
Best Local Similarity 24.9%; Pred. No. 0.0011;
Matches 44; Conservative 31; Mismatches 74; Indels 28; Gaps 4;
QY 3 ILLVGIWHLGIVMLPRLSHRLKGYRTOTISYNLSALDDAIRRLDRSLTHSPNAV 62
DB 1 MLVGIWMTAWMLPLLRPMKANGALPALPGYVSLGPPQAVPHILIRASRVQVI 60
QY 63 GHSGLGLVIRKYLRSRAPSCEITLSHVAIGSPLOGASIVNKIEQLGVALGNSA---EF 119
DB 61 GHSGLGLVIRKYLRSRAPSCEITLSHVAIGSPLOGASIVNKIEQLGVALGNSA---EF 118
QY 120 GLKHEHDSRRYPOKSGS--IAGTTPGLRSLRLDPLDSGTVVTEETKAGTDTIAT 176
DB 119 GPARDWDGAEITGQVAGNVA-----LAETRLPLGLRDHCIV 152

```

```

RESULT 9
ID 055748 PRELIMINARY; PRT; 249 AA.
AC 055748;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein slr0421.
GN slr0421.
OS Synecocystis sp. (Strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 6803;
RA Tabata S.;
RA Submitted (Aug-1995) to the EMBL/Genbank/DBJ databases.

```

```

RESULT 10
ID 09PHH3 PRELIMINARY; PRT; 251 AA.
AC 09PHH3;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein Xfa0032.
GN XFA0032.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=945C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.U.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarengas R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carrer D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Gardner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes S.A., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,

```

```

[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64k to 92k of the genome."
RL DNA Res. 2:153-166(1995).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugita M., Sasaoka S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Ntaro K., Okumura S.,
RA Shimizu S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL; D64002; BAA10393.1; -.
DR PIR; S76547; S76547.
DR GO; GO:0003824; F: catalytic activity; IEA.
DR InterPro; IPR000379; Ser_pstc.
KM Hypothetical protein, Complete proteome.
SQ SEQUENCE 249 AA; 27534 MM; F36D75D573232008 CRC64;
Query Match 13.0%; Score 121.5; DB 16; Length 249;
Best Local Similarity 32.7%; Pred. No. 0.004;
Matches 37; Conservative 20; Mismatches 39; Indels 17; Gaps 5;
QY 60 ALVGHSLGLVIRKYLRSRAPS--CEITLSHVAIGSPLOGASIVNKIEQLGVALGNSA 118
DB 72 SVVGFPMGGILWVEVLS--SRNPWLELDLVLVIGSPLOGASIVNKIEQLGVALGNSA 128
QY 119 GLKHEHDSRRYPOKSGS--IAGTTPGLRSLRLDPLDSGTVVTEETKAGTDTIAT 169
DB 129 LGRRROYLAKRTQKIPVLIVAGSVNGS-----GDVPIPRESTRING 170

```

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Neto L.E.S.,  
 RA Nhari A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pequeiro J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawaaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA de Souza A.P., Terenzi M.F., Trufil D., Tsai S.M., Tsubako M.H.,  
 RA Valada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zazo M.A., Zatz M., Zeldanis J., Zetbal J.C.,  
 RA "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
 RL Nature 406:151-159(2000).  
 DR EMBL: AE003851; AAF85601.1; -  
 DR PIR: H82864; H82864.  
 DR GO: GO:0003824; F: catalytic activity; IEA.  
 DR InterPro: IPR008262; Lipase\_AS.  
 DR InterPro: IPR000379; Ser esterase.  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 251 AA; 27097 MW; 0029A52776B07054 CRC64;

Query Match 12.0%; Score 112.5; DB 16; Length 251;  
 Best Local Similarity 24.7%; Pred. No. 0.028;  
 Matches 48; Conservative 23; Mismatches 78; Indels 45; Gaps 8;

QY 3 IIVHGLVYHGLVWHPHLSRLHKLGYRTQTSYNSLAI--DDEALFPRDLRSLTHASPN 60  
 DB 26 IIVHGLVYHGLVWHPHLSRLHKLGYRTQTSYNSLAI--DDEALFPRDLRSLTHASPN 82  
 QY 61 LVGHSGLGLVIRKYLESPAPSCETLSHVAIG--SPLOASIVYKIKIQLGLVALGNSAE 118  
 DB 83 LVGHSYGAVI-----TEAGDERVKALVYIAFAPSGESVA-----120  
 QY 119 FGLKEHDESRYPQKSG-----SLAGTIPGLASLLIRD-PLSDGTVYEETKIA 168  
 DB 121 -----DINKKYPFPPSGYNHLSDXKEGFLMTPFGEVKYLAQDIPLEQRLIATQHPIR 174  
 QY 169 GMTDHAISTSYE 182  
 DB 175 GANFEKYSAAAME 188

RESULT 11  
 Q8KUJ3 PRELIMINARY; PRT; 254 AA.

ID 08KUJ3  
 AC 08KUJ3  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DB Putative lipase LipB.  
 DB Legionella pneumophila.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
 CC Legionellaceae; Legionella.  
 CX NCBI\_TaxID=446;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=130b;  
 RX MEDLINE=22095963; PubMed=12101309;  
 RA Aragon V., Rosier O., Cianciotto N.P.;  
 RT "Legionella pneumophila genes that encode lipase and phospholipase C  
 RT activites";  
 RL Microbiology 148:2223-2231(2002).  
 DR EMBL: AF454864; AAM73853.1; -  
 DR GO: GO:0003824; F: catalytic activity; IEA.  
 DR InterPro: IPR000379; Ser esterase.  
 SQ SEQUENCE 254 AA; 28850 MW; 999C2E079C102011 CRC64;

Query Match 11.9%; Score 111.5; DB 2; Length 254;  
 Best Local Similarity 24.4%; Pred. No. 0.035;  
 Matches 47; Conservative 30; Mismatches 81; Indels 35; Gaps 8;

QY 3 IIVHGLVYHGLVWHPHLSRLHKLGYRTQTSYNSLAI--IDDEALFPRDLRSLTHASPN 59  
 DB 57 IIVHGLVYHGLVWHPHLSRLHKLGYRTQTSYNSLAI--IDDEALFPRDLRSLTHASPN 114  
 QY 60 -----LVGHSGLGLVIRKYLESPAPSCETLSHVAIG--SPLOASIVYKIKIQLGLVALGNSAE 109  
 DB 115 PAIKHIFVTHSGIIVRALSKL--SQKQLHVGVLIMLAPNGSKAKSLKIFPMI 172  
 QY 110 GVALGNSAFGLKEHDESRYPQKSG-----SLAGTIPGLASLLIRD-PLSDGTVYEETK 166  
 DB 173 SYFKIPKALSSGQTSYVARVPDIPKQGIAGRF-----DAKVPDYAR 217  
 QY 167 IAGMTDHAISTT 179  
 DB 218 LQGTQEVLIINT 230

RESULT 12  
 Q98KT5 PRELIMINARY; PRT; 256 AA.

ID 098KT5  
 AC 098KT5  
 DT 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DB Hypothetical protein m11329.  
 GN M11329.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Rhizobiaceae; Mesorhizobium.  
 CX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF30309;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu B., Kato T., Sasamoto S.,  
 RA Matsumoto M., Ideesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AF002997; BAB46729.1; -  
 DR GO: GO:0003824; F: catalytic activity; IEA.  
 DR InterPro: IPR000379; Ser esterase.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 256 AA; 26643 MW; 7AC63AB6830E716C CRC64;

Query Match 11.4%; Score 107; DB 16; Length 256;  
 Best Local Similarity 24.6%; Pred. No. 0.035;  
 Matches 42; Conservative 28; Mismatches 67; Indels 34; Gaps 6;

QY 3 IIVHGLVYHGLVWHPHLSRLHKLGYRTQTSYNSLAI--IDDEALFPRDLRSLTHASPN 60  
 DB 35 IIVHGLVYHGLVWHPHLSRLHKLGYRTQTSYNSLAI--IDDEALFPRDLRSLTHASPN 91  
 QY 61 LVGHSGLGLVIRKYLESPAPSCETLSHVAIG--SPLOASIVYKIKIQLGLVALGNSAE 108  
 DB 92 LVGHSYGAVI-----TEAGDERVKALVYIAFAPSGESVA-----120  
 QY 109 LCVAGNSAFGLKEHDESRYPQKSG-----SLAGTIPGLASLLIRD-PLSDGTVYEETK 159  
 DB 152 LVI-----DPAKVPF---SAVADLPFAIESLALNHQDPLINHT 185

RESULT 13  
 Q98JB5 PRELIMINARY; PRT; 233 AA.

ID 098JB5  
 AC 098JB5  
 DT 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DB Hypothetical protein m112018.

GN ML2018.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 ON NCBI\_TaxID=381;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214969;  
 RA Kato T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasaoka S.,  
 Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,  
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,  
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
 Takeuchi C., Yamada M., Tabata S.  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti."  
 RL DNA Res. 7:331-338 (2000).  
 DR EMBL; AP002998; BAB49251.1;  
 DR GO; GO:0003824; P: catalytic activity; IEA.  
 DR InterPro; IPR008262; Lipase AS.  
 DR InterPro; IPR000379; Ser esters.  
 DR PROSITE; PS00120; LIPASE\_SER; 1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 233 AA; 23925 MW; F93FBE505C548D63 CRC64;

Query Match 11.3%; Score 105.5; DB 16; Length 233;  
 Best Local Similarity 35.1%; Pred. No. 0.11;  
 Matches 40; Conservative 20; Mismatches 31; Indels 23; Gaps 8;  
 Db 3 IILVHGSLGVLVMEPLSHRLKLGRTQTISY--SLAIDDEAIFRRLDRSLTHASPN 59  
 9 VVLVHGSLGVLVMEPLSHRLKLGRTQTISY--SLAIDDEAIFRRLDRSLTHASPN 63  
 Qy 60 A-LVHGSLGVLVMEPLSHRLKLGRTQTISY--SLAIDDEAIFRRLDRSLTHASPN 100  
 64 VILVHGSLGVLVMEPLSHRLKLGRTQTISY--SLAIDDEAIFRRLDRSLTHASPN 115

RESULT 14  
 ID O8UKB1 PRELIMINARY; PRT; 268 AA.  
 AC O8UKB1;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein Atus212.  
 GN Atus212 OR AGR PAT 295.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OG Plasmid AT.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 ON NCBI\_TaxID=176299;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=2168550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,  
 Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Beeve D. Sr.,  
 Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,  
 Kutyavin T., Levy R., Li M.-J., McClelland E., Palmer A.,  
 Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 Zhang S., Yeo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 Rastumay E.W., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 Neeter E.W.,  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 C58."  
 RL Science 294:2317-2323 (2001).  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2168551; PubMed=11743194; Miller N., Blanchard M.,  
 Goodner B., Hinkle G., Gattung S., Weller N., Gattung S.,  
 Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,

RA Houtmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
 RA Molam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gerson J., Lomo C., Sear C., Strub G.,  
 RA Flanagan C., Slater S.,  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 Agrobacterium tumefaciens C58."  
 RL Science 294:2323-2328 (2001).  
 DR EMBL; AB008944; AL45901.1; ALT\_INT.  
 DR EMBL; AB007891; AK90583.1;  
 DR PIR; AG3185; AG3185.  
 DR GO; GO:0046821; C: extrachromosomal DNA; IEA.  
 DR GO; GO:0003824; P: catalytic activity; IEA.  
 DR InterPro; IPR008262; Lipase AS.  
 DR InterPro; IPR000379; Ser esters.  
 DR PROSITE; PS00120; LIPASE\_SER; 1.  
 KM Hypothetical protein; Plasmid; Complete proteome.  
 SQ SEQUENCE 288 AA; 30807 MW; 0C5B6AA4578FEAB CRC64;

Query Match 11.3%; Score 105.5; DB 16; Length 288;  
 Best Local Similarity 28.3%; Pred. No. 0.15;  
 Matches 45; Conservative 19; Mismatches 68; Indels 27; Gaps 6;  
 Db 3 IILVHGSLGVLVMEPLSHRLKLGRTQTIS--VNSLAIDDEAIFRRLDRSLTHASPN 60  
 63 VVLVHGSLGVLVMEPLSHRLKLGRTQTIS--VNSLAIDDEAIFRRLDRSLTHASPN 119  
 Qy 61 LVHGSLGVLVMEPLSHRLKLGRTQTIS--VNSLAIDDEAIFRRLDRSLTHASPN 118  
 120 LVHGSLGVLVMEPLSHRLKLGRTQTIS--VNSLAIDDEAIFRRLDRSLTHASPN 168  
 Qy 119 FGLKHHDESRYPQKSGSLGRTITPLGLRLSLRLDPLSD 157  
 169 FVLETTDGPXY-----VSPAKKSGFADVDSD 198

RESULT 15  
 ID Q92XY1 PRELIMINARY; PRT; 243 AA.  
 AC Q92XY1;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein RAL107.  
 GN RAL107 OR SMA2023.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OG Plasmid pSymb (megaplasmid 1).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
 ON NCBI\_TaxID=382;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=2136509; PubMed=11461432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 Barlow-Hubler F., Bowser L., Capela D., Gallibert F., Gouzy J.,  
 Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn M.L.,  
 Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 Yeh K.-C., Davis R.W., Federle N.A., Long S.R.,  
 RT "Nucleotide sequence and predicted functions of the entire  
 Sinorhizobium meliloti pSymb megaplasmid."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).  
 DR EMBL; AB007296; AK65765.1;  
 DR PIR; C95400; C95400.  
 DR GO; GO:0046821; C: extrachromosomal DNA; IEA.  
 DR GO; GO:0003824; P: catalytic activity; IEA.  
 DR InterPro; IPR008262; Lipase AS.  
 DR InterPro; IPR000379; Ser esters.  
 DR PROSITE; PS00120; LIPASE\_SER; 1.  
 KM Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 243 AA; 25940 MW; 59A484BD1DBA850 CRC64;

Query Match 11.2%; Score 104.5; DB 16; Length 243;  
 Best Local Similarity 28.9%; Pred. No. 0.15;

Matches 55; Conservative 19; Mismatches 77; Indels 39; Gaps 10;

```
QY      3 IILVHGLMHHGIVMHPISHRLHLKGYRQTISYNSIAI--DDEAIIPRR-IDRSLTHASPN 59
      18 VVLVHGAFADSGGKGVYDNLTKRGYRV-TIVQNPLTSLDDVAATRPALER---QDGPV 73
QY      60 ALVGHSLGGLVTKRYLBSRAPSCETLSHVVAIGSPLOGASIVNKIEQLGLGVALGNSAEF 119
      74 ILVGHSWGQGVITE--TGIDPKVAGLVVVSAL-SPDAGETTAQOYE-----GFAPAAEF 124
QY      120 GLKEHDDSRYPQKSGSIAGTIFPLGLRSLLRDPIDSDGTIVTEETKIAGMTD----HI 174
      125 VLETTKDGFGY-----VSPAKFKAGFAHDVSDAD-----VAFWRDAQVPIM 166
Db
QY      175 AISTTSYENA 184
      167 SAFATKLENA 176
```

Search completed: April 29, 2004, 11:36:59  
Job time : 62 secs



Matches 51; Conservative 31; Mismatches 79; Indels 46; Gaps 11;

QY 3 IIVHGL-----YHGLVHPLSLRLKLGRTTISYNLAIDDEA---IFRR 48  
DB 11 IIVHGLFGRDRIGSHHYFHG-----IKQALNECG-ASVFVPIISAANDNEARGDQLKQ 64  
QY 49 LD--RSLTASPNALVGHSLGGLVIRKYESRAPSCETISHVAIGSPLOGASIVNKIE- 105  
DB 65 IHNLRQVGAQRVNLIGHSGGALT-ARYVAALAP--ELIASYTSVSGPNHSGELADRLRL 121  
QY 106 -----QGLGVALGNASAEFG--LKEHDDSRYPQKSGSIAGTIPGLRLRLDPDSDG 158  
DB 122 AFVPGRLGETVVAALTTSPSAFLSALSGHPLPQNA-----LNAL---NALTTDG 168  
QY 159 TVTVEETKLAGMTDHAISTSYENAV 185  
DB 169 VAAFNROYPOGLPDRWGGMGPACVNAV 195

RESULT 2  
US-08-360-758-3  
Sequence 3, Application US/08360758  
Patent No. 6074863  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
APPLICANT: Patkar, Shankant A  
APPLICANT: Egel-Mittani, Michi  
APPLICANT: Borich, Kim  
APPLICANT: Clausen, Ib G  
APPLICANT: Hansen, Mogens T  
TITLE OF INVENTION: C. ANTARCTICA LIPASE AND LIPASE VARIANTS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESS: No. 6074863 No. 6074863disk of No. 6074863th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Tape  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,758  
FILING DATE: 22-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK PCT/DK93/00225  
FILING DATE: 03-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 277 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-360-758-3

Query Match 10.4%; Score 97; DB 3; Length 277;  
Best Local Similarity 24.6%; Pred. No. 0.015;  
Matches 51; Conservative 31; Mismatches 79; Indels 46; Gaps 11;  
QY 3 IIVHGL-----YHGLVHPLSLRLKLGRTTISYNLAIDDEA---IFRR 48

DB 11 IIVHGLFGRDRIGSHHYFHG-----IKQALNECG-ASVFVPIISAANDNEARGDQLKQ 64  
QY 49 LD--RSLTASPNALVGHSLGGLVIRKYESRAPSCETISHVAIGSPLOGASIVNKIE- 105  
DB 65 IHNLRQVGAQRVNLIGHSGGALT-ARYVAALAP--ELIASYTSVSGPNHSGELADRLRL 121  
QY 106 -----QGLGVALGNASAEFG--LKEHDDSRYPQKSGSIAGTIPGLRLRLDPDSDG 158  
DB 122 AFVPGRLGETVVAALTTSPSAFLSALSGHPLPQNA-----LNAL---NALTTDG 168  
QY 159 TVTVEETKLAGMTDHAISTSYENAV 185  
DB 169 VAAFNROYPOGLPDRWGGMGPACVNAV 195

RESULT 3  
US-09-252-991A-27337  
Sequence 27337, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,788  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 27337  
LENGTH: 351  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27337

Query Match 10.3%; Score 96.5; DB 4; Length 351;  
Best Local Similarity 27.7%; Pred. No. 0.023;  
Matches 31; Conservative 23; Mismatches 45; Indels 13; Gaps 5;

QY 3 IIVHGLVHGLV-----HPLSLRLKLG---YRTQTISYNLAIDDEAIFRLDR--S 52  
DB 76 IIVHGLFGRKSVGPVDYHAIIVPALEKQAKFASQSVNSNVRGEGLAQVEVLA 135  
QY 53 LTHASPNALVGHSLGGLVIRKYESRAPSCETISHVAIGSPLOGASIVNKI 104  
DB 136 LTGAEKVNLIGHSGGWTV-RVYAGVAP--QLVASVTMGTPHKGTPVADAV 184

RESULT 4  
US-09-328-352-5069  
Sequence 5069, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-039A  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5069  
LENGTH: 453  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5069

Query Match 10.3%; Score 96.5; DB 4; Length 453;  
Best Local Similarity 29.5%; Pred. No. 0.034;  
Matches 39; Conservative 18; Mismatches 42; Indels 33; Gaps 7;  
QY 3 IIVHGLVHGLV-----GLVHPLSLRLKLG---LGYRT-OTISYNLAIDDEAIFR 47



Db 164 VILVHGLGMNHLTWSNHHYGGIGERLLAQRHNMVLNTVTGRISANRS-----FS 217  
QY 48 RLDSRLTASANA---LVHSLGLGVTKRYL---ESRAPSCETLSHVVALGSPLOGAS 99  
Db 218 NLBLDLVORNPRTISIDILGHSMGLVSRSLFYGKQNMVQIMHVENLVICISPHRGA- 276  
QY 100 IVNKLKQGLGV 111  
Db 277 ---VLERGFAL 285

RESULT 5  
US-09-107-532A-4594  
Sequence 4594, Application US/09107532A  
Patent No. 6583275

## GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

## COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

## ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277

## INFORMATION FOR SEQ ID NO: 4594:

## SEQUENCE CHARACTERISTICS:

LENGTH: 316 amino acids  
TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

## FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...316

SEQUENCE DESCRIPTION: SEQ ID NO: 4594:

US-09-107-532A-4594

Query Match 10.3%; Score 96; DB 4; Length 316;  
Best Local Similarity 31.2%; Pred. No. 0.023; Mismatches 32; Indels 28; Gaps 6;  
Matches 35; Conservative 17;

QY 61 LVHSLGLGVTKRYL---ESRAPSCETLSHVVALGSPLOGAS 112  
Db 170 LVHSLGLGVTKRYL---ESRAPSCETLSHVVALGSPLOGAS 112  
170 LVHSLGLGVTKRYL---ESRAPSCETLSHVVALGSPLOGAS 112  
170 LVHSLGLGVTKRYL---ESRAPSCETLSHVVALGSPLOGAS 112

QY 113 LGNSAEFGLEKHEHDESRYPQKSGSIAGTIFLGRLSLRLDPLD---SDGTIV 160  
US-08-449-015-10

Db 227 VVS-----NRY-QDYQWIMGNIPITRREFLLAGQDERDLSGDTV 265

## RESULT 6

US-08-034-650-10  
Sequence 10, Application US/08034650  
Patent No. 5641671

## GENERAL INFORMATION:

APPLICANT: BOS, Janet W.  
APPLICANT: BRENNEN, Leon G.  
APPLICANT: VERBES, Cornelis T.

APPLICANT: VISSER, Christiaan

TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE

TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

STREET: 1615 L. Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036-5601

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/034,650

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/727,235

FILING DATE: 03-JUL-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kokulis, Paul N.

REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: PAK/5970/91731

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3000

TELEFAX: (202) 822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 358 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-034-650-10

Query Match 10.3%; Score 96; DB 1; Length 358;  
Best Local Similarity 21.2%; Pred. No. 0.027; Mismatches 65; Indels 62; Gaps 8;  
Matches 43; Conservative 33;

QY 3 IIVHGL-----YHGLVHHPPLSHRLKGYRTOTISYNSLAID-----EAI 45  
Db 50 VILVHGLAGTDKFRANVVDYWGIGSDLSH-----GAKYVYANLSFGSDGPNRGEQL 104  
QY 46 FRLLDRSL--THASPNALVHSLGLVTKRYLESRAPSCETLSHVVALGSPLOGAS 103  
Db 105 LAYKQVLATGATKYNLIGHSGGGIT--SRVYAAVAF--QVYASVTTTTPHRSSEFADF 161  
QY 104 IEQGLGVALLGNSAEFGLEKHEHDESRYPQKSGSIAGTIFLGRLSLRLDPLDSDGTIV 160  
Db 162 VQ-----DVLKTDPTGLSSTVIAAFVAVFGTLVSS 191

QY 161 --TVEETKLAGMTDHAISTTSY 181  
Db 192 SHNTQDAAALATLTITTAQTATY 214

RESULT 7  
US-08-449-015-10

Sequence 10, Application US/08449015  
Patent No. 5804409  
GENERAL INFORMATION:  
APPLICANT: BOS, Janette W.  
APPLICANT: FRENKEN, Leon G.  
APPLICANT: VERKIPS, Cornelis T.  
APPLICANT: VISSER, Christiaan  
TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE  
TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1615 L. Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,015  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/727,235  
FILING DATE: 03-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kokulis, Paul N.  
REGISTRATION NUMBER: 16, 773  
REFERENCE/DOCKET NUMBER: PNK/5970/91731  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 358 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-449-015-10

Query Match 10.3%; Score 96; DB 1; Length 358;  
Best Local Similarity 21.2%; Pred. No. 0.027;  
Matches 43; Conservative 33; Mismatches 65; Indels 62; Gaps 8;

QY 3 IIVHGL-----YMGVLVHPLSHRLKLGRTQTISVNSLAID-----EAI 45  
DB 50 VILVHGLAGTDKFRANVVDYWGQISDLQSH-----GAKYVYANLGGFOSDDGPNRGEO 104  
QY 46 FRILDSL--THASPNALVGHSLGIVIRYLESRAPSCETLSHVVAIGSPLOGASTYNK 103  
DB 105 LAVVKQVLATGATKYNLIGSGQGLT--SRVAAVAP--QVASTYTTIGTPHRSGEFADF 161  
QY 104 IEQLGVALGNSEFGLKEHDESRYPKSGSIAGTIPLGRLSLDPLDSDGTV--- 160  
DB 162 VQ-----DVAKTDPTGLSTVIAAFVVFGLVSS 191  
QY 161 --TWEEKTAGMTDAISTSY 181  
DB 192 SHNTODALAAHRTLTAAQTATY 214

RESULT 8  
US-09-543-681A-6874  
Sequence 6874, Application US/09543681A  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 6874  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-6874

Query Match 9.9%; Score 92.5; DB 4; Length 333;  
Best Local Similarity 23.8%; Pred. No. 0.06;  
Matches 49; Conservative 27; Mismatches 67; Indels 63; Gaps 11;

QY 2 QIIVHGL-----YMGVL-----VMEPLSHRLKLGIR-----TQITSY 35  
DB 68 RLVIFFGLGNGNPSYPVHGMLAAKARGLGVMM-----FRGSGEPNRQKRLY 117  
QY 36 NSLAIDDEAIFRILDSLTHASPNALVGHSLGIVIRYLESRAPSCETLSHVVAIGSP 95  
DB 118 HSEETDARYFLNMLKRGEOPTAVGYSLGKMLAYTLASGERA-VLDAAYIVSAPL 176  
QY 96 OGASIVNKLIEQLGVALGNSEF-----GLKEHDES--RYPKSGSIAGTIPLGLRS 147  
DB 177 MLEPCSTKIER-----GFSRFYQWYLLKGLKRNTRKLIHYPC-----SLPISL 222  
QY 148 L-LRDPDSGDTVYVEETKIAGMTD 172  
DB 223 IKSIKRLRPFDDILT--AKIHGFKD 245

RESULT 9  
US-08-602-359A-38  
Sequence 38, Application US/08602359A  
Patent No. 5942430  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAPPIA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald V.  
APPLICANT: WARREN, Patrick V.  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & RICHARDSON P.C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,359A  
FILING DATE: February 16, 1996  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 11:37:05 ; Search time 43 Seconds  
(without alignments)  
1192.560 Million cell updates/sec

Title: US-10-603-260-5  
Sequence: 1 MQLIVHGLVYMGVWHPPLS.....KINQMTDHAISTYENAV 185

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues  
Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubppaa/PCRNEM\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubppaa/PCRNUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubppaa/US09C\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	936	100.0	185	US-10-603-260-5	Sequence 5, Appl1
2	110	11.8	249	US-10-359-493-8213	Sequence 8213, Ap
3	97	10.4	447	US-10-421-654-16	Sequence 16, Appl
4	91.5	9.8	262	US-09-903-410-38	Sequence 38, Appl1
5	91.5	9.8	262	US-10-027-805-38	Sequence 38, Appl1
6	91.5	9.8	262	US-10-027-804-38	Sequence 38, Appl1
7	90.5	9.7	286	US-10-156-761-10998	Sequence 10998, A
8	90.5	9.7	262	US-10-147-467-2	Sequence 2, Appl1
9	90	9.6	443	US-10-434-599-24654	Sequence 24654, A
10	88.5	9.5	518	US-10-038-174-1	Sequence 1, Appl1
11	88.5	9.5	627	US-10-282-122A-70475	Sequence 70475, A
12	87.5	9.3	286	US-10-156-761-14086	Sequence 14086, A
13	87.5	9.3	286	US-10-282-122A-47438	Sequence 47438, A
14	87	9.3	345	US-10-282-122A-76642	Sequence 76642, A
15	87	9.3	505	US-09-938-803-20	Sequence 20, Appl

16	85	9.1	1216	US-10-369-493-1505	Sequence 1505, Ap
17	82.5	8.8	275	US-10-369-493-4463	Sequence 4463, Ap
18	82.5	8.8	275	US-10-369-493-7223	Sequence 7223, Ap
19	82.5	8.8	279	US-10-282-122A-62627	Sequence 62627, A
20	82.5	8.8	301	US-10-282-122A-49252	Sequence 49252, A
21	82.5	8.8	323	US-10-282-122A-64350	Sequence 64350, A
22	82.5	8.8	432	US-10-156-761-13452	Sequence 13452, A
23	82.5	8.8	470	US-10-421-654-58	Sequence 58, Appl
24	82	8.8	244	US-10-289-762-560	Sequence 360, App
25	81.5	8.7	259	US-10-369-493-21672	Sequence 21672, A
26	81	8.7	332	US-10-050-882-99	Sequence 99, Appl1
27	80.5	8.6	244	US-10-369-493-19863	Sequence 19863, Ap
28	80.5	8.6	252	US-10-369-493-16706	Sequence 16706, A
29	80.5	8.6	259	US-10-369-493-16706	Sequence 16706, A
30	80.5	8.6	267	US-10-282-122A-46834	Sequence 46834, A
31	80.5	8.6	301	US-10-156-761-786	Sequence 786, Ap
32	80.5	8.6	331	US-09-815-242-10799	Sequence 10799, A
33	80.5	8.6	331	US-10-282-122A-57052	Sequence 57052, A
34	80.5	8.6	360	US-10-424-599-221830	Sequence 221830, A
35	80.5	8.6	361	US-09-896-908-2	Sequence 2, Appl1
36	80.5	8.6	1460	US-10-282-122A-55255	Sequence 55255, A
37	80	8.5	225	US-10-425-114-63517	Sequence 63517, A
38	80	8.5	240	US-10-425-114-45130	Sequence 45130, A
39	80	8.5	257	US-10-282-122A-77391	Sequence 77391, A
40	80	8.5	587	US-10-425-114-68123	Sequence 68123, A
41	80	8.5	1129	US-10-156-761-8946	Sequence 8946, Ap
42	79.5	8.5	302	US-10-282-122A-47335	Sequence 47335, A
43	79.5	8.5	303	US-10-282-122A-50776	Sequence 50776, A
44	79.5	8.5	403	US-10-156-761-14428	Sequence 14428, A
45	79	8.4	270	US-10-206-576-188	Sequence 188, App

## ALIGNMENTS

RESULT 1  
US-10-603-260-5  
; Sequence 5, Application US/10603260  
; Publication No. US20040009570A1  
; GENERAL INFORMATION:  
; APPLICANT: CJ Corporation  
; TITLE OF INVENTION: An alkaline lipase from *Vibrio metschnikovii* RH530 and a  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/603,260  
; CURRENT FILING DATE: 2003-06-24  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Koparentin 1.71  
; SEQ ID NO 5  
; LENGTH: 185  
; TYPE: PRT  
; ORGANISM: *Vibrio metschnikovii* RH530  
US-10-603-260-5

Query Match	100.0%	Score 936;	DB 15;	Length 185;
Best Local Similarity	100.0%	Pred. No. 4.1e-96;		
Matches 185;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MQLIVHGLVYMGVWHPPLSRLHLKLGRTQTISTNSLAIDDEAIFRRDRSLTASPNV	60	
DB	1	MQLIVHGLVYMGVWHPPLSRLHLKLGRTQTISTNSLAIDDEAIFRRDRSLTASPNV	60	
QY	61	LVHSGIGLVYKRYLSEKRAPSCETTSHTVVAIGSPLOGASIVNKIKRLGIGVALGNSAERG	120	
DB	61	LVHSGIGLVYKRYLSEKRAPSCETTSHTVVAIGSPLOGASIVNKIKRLGIGVALGNSAERG	120	
QY	121	LKEHDESRYPKSGSIACTIPGLKRLSLRPLSDGTVTVEETKIAQMTDHAISTTS	180	
DB	121	LKEHDESRYPKSGSIACTIPGLKRLSLRPLSDGTVTVEETKIAQMTDHAISTTS	180	
QY	181	YENAV 185		
DB	181	YENAV 185		





APPLICANT: ROBERTSON, Daniel E.  
MURPHY, Dennis  
REID, John  
MAFFIA, Anthony  
LINK, Steven  
SWANSON, Ronald V.  
WARREN, Patrick V.  
KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSER: FISH & RICHARDSON P. C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/027,805  
FILING DATE: 21-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,359  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 262 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 38:  
US-10-027-805-38  
Query Match 9.8%; Score 91.5; DB 13; Length 262;  
Best Local Similarity 24.6%; Pred. No. 0.14;  
Matches 46; Conservative 16; Mismatches 58; Indels 67; Gaps 7;  
QY 3 IILVHGIVMGVLMPLSHRLHKLGYRTQTISY-----NSLAIDDEAIF 46  
DB 16 VLVHVGIGHSRGYGRILIKELNAYAGGYTFDWPFGHKSPGKRGHTSVEMAEIIDSIIIE 75  
QY 47 RLDRSLTHASPNALVGHSLGIVIRYLESR-----APSCETLSHVAIG 92  
DB 76 EIREKEPF-----LFGSLGGLTVIRYAETRPDKIRGLIASPPALAKSPETGFWVAL- 127  
QY 93 SPLQASIVNKIKIQGLGVALGN--SAEFGLEKHDDSESYPOKSSGIAGTIPLGLRSILL 150  
DB 128 -----AKPIGKIAP---GVVLSNGIKPELISRNRDAVRY-----V 160  
QY 151 RDPPLDSD 157  
DB 161 EDPPLVHD 167  
RESULT 5  
US-10-027-804-38  
Sequence 38, Application US/10027804  
Publication No. US20030054530A1  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
MURPHY, Dennis

REID, John  
MAFFIA, Anthony  
LINK, Steven  
SWANSON, Ronald V.  
WARREN, Patrick V.  
KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSER: FISH & RICHARDSON P. C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/027,804  
FILING DATE: 21-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,359  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 262 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 38:  
US-10-027-804-38  
Query Match 9.8%; Score 91.5; DB 14; Length 262;  
Best Local Similarity 24.6%; Pred. No. 0.14;  
Matches 46; Conservative 16; Mismatches 58; Indels 67; Gaps 7;  
QY 3 IILVHGIVMGVLMPLSHRLHKLGYRTQTISY-----NSLAIDDEAIF 46  
DB 16 VLVHVGIGHSRGYGRILIKELNAYAGGYTFDWPFGHKSPGKRGHTSVEMAEIIDSIIIE 75  
QY 47 RLDRSLTHASPNALVGHSLGIVIRYLESR-----APSCETLSHVAIG 92  
DB 76 EIREKEPF-----LFGSLGGLTVIRYAETRPDKIRGLIASPPALAKSPETGFWVAL- 127  
QY 93 SPLQASIVNKIKIQGLGVALGN--SAEFGLEKHDDSESYPOKSSGIAGTIPLGLRSILL 150  
DB 128 -----AKPIGKIAP---GVVLSNGIKPELISRNRDAVRY-----V 160  
QY 151 RDPPLDSD 157  
DB 161 EDPPLVHD 167  
RESULT 7  
US-10-156-761-10998  
Sequence 10998, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMTBA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI

```
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10998
LENGTH: 286
TYPE: PR1
ORGANISM: Streptomyces avermitilis
US-10-156-761-10998
```

```
Query Match
Best Local Similarity 20.9%; Score 90.5; DB 14; Length 286;
Matches 44; Conservative 38; Mismatches 84; Indels 45; Gaps 8;
```

```
QY 3 IIVHGLVHGLVMEPLSHRLHKLQ-YRTQTISNSLAIDDEAIFRRIDRSIT-----54
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 77 VVILHGFINDRSYFVLLRRSLTQHGROQIESLNSPLTCDITRAELLGRHIEVICERTG 136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 55 HASPNALVGHSLGGVIVKRYLESRAPCSTLSHVAIGSP-----L 95
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 137 HAIVD-IVGHSIGGLIARIYVQCLGGDIRVRI-LVTLGPHSGTRVAPLANHPVRQMR 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 96 OGASIVNKIEQLGLVA-----LGNSEFGLKEHDESRYPQKSGSIAGTIP 143
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 195 PGSEVIEELQRPAPGCRTHFVSFMSDDLHMLDPLEACVCEHDLVAVQVRVSGI-GHIAL 253
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 144 GLRSLL---LRPDLSDGTYVEETKIAGMT 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 254 PVHPAVATGIRKALDSEGTGAAPRTDGLT 284
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 8  
US-10-147-467-2

```
Sequence 2, Application US/10147467
Publication No. US20030027295A1
GENERAL INFORMATION:
APPLICANT: TAKESHI, Ishii
APPLICANT: SATOSHI, Matsuda
TITLE OF INVENTION: ESTERASE GENE AND ITS USE
FILE REFERENCE: 20-43346P
CURRENT APPLICATION NUMBER: US/10/147,467
CURRENT FILING DATE: 2002-05-17
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 363
TYPE: PR1
ORGANISM: Burkholderia cepacia
US-10-147-467-2
```

```
Query Match
Best Local Similarity 9.7%; Score 90.5; DB 14; Length 363;
Matches 47; Conservative 20; Mismatches 57; Indels 51; Gaps 8;
```

```
QY 3 IIVHGLVHGLVMEPLSHRLHKLQYRTQTISNSLAIDDEAIFRR 48
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 55 IIVHGLVHGLVMEPLSHRLHKLQYRTQTISNSLAIDDEAIFRR 112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 49 IDRSL--THASPNALVGHSLGGVIVKRYLESRAPCSTLSHVAIGSPLOGASIVNKIEQ 106
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 113 VKQVLAATGAQKNLIGHSGGGLT-SRYVASVAP--ELVASVTTSTPHWGSQFADPVQO 169
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 107 LGLVGLVHGLVMEPLSHRLHKLQYRTQTISNSLAIDDEAIFRR 161
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 170 L-----LQTDPTGLSTVLAFAFALGTLT 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 9  
US-10-424-599-246354

```
Sequence 246354, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic, David K
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 246354
LENGTH: 443
TYPE: PR1
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_6448C.1.pep
US-10-424-599-246354
```

```
Query Match
Best Local Similarity 9.6%; Score 90; DB 12; Length 443;
Matches 43; Conservative 21; Mismatches 40; Indels 78; Gaps 9;
```

```
QY 9 LNHGLVME-----PLSHRLHKLQYRTQTISNSLAIDDEAIF-----46
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 131 LYNPLRKHITGQWAPVYSLQDLGY-----AGETLFGAPYDFRYGLAEGHP 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 47 -----RRDRSLTHAS-----PNALVGHSLGGVIVKRYLESRAPC--ETLSHVA 90
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 180 SQVSKFELKDLKILBANSNNGKRVILSHSLGSLFVLQLNRPPEMRKKTFFHFLA 239
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 91 IGSPLOGASIVNKIEQLGLVGLVHGLVMEPLSHRLHKLQYRTQTISNSLAIDDEAIF 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 240 LSAPWGA-----IDEM-----YTFASGNTLG-VPL-VDPFLV 270
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 151 RD 152
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 271 RD 272
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 10  
US-10-098-174-1

```
Sequence 1, Application US/10098174
Publication No. US20020169288A1
GENERAL INFORMATION:
APPLICANT: HOOK, Magnus
APPLICANT: BOWDEN, Maria
TITLE OF INVENTION: COLLAGEN-BINDING ADHESIN FROM STAPHYLOCOCCUS EPIDERMIDIS.
FILE REFERENCE: P07136US01/BAS
CURRENT APPLICATION NUMBER: US/10/098,174
CURRENT FILING DATE: 2002-03-15
PRIOR APPLICATION NUMBER: 60/275,718
PRIOR FILING DATE: 2001-03-15
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 518
TYPE: PR1
ORGANISM: Staphylococcus epidermidis
US-10-098-174-1
```

```
Query Match
Best Local Similarity 9.5%; Score 88.5; DB 13; Length 518;
Matches 47; Conservative 32; Mismatches 55; Indels 89; Gaps 12;
```

```
QY 3 IIVHGLVHGLVMEPLSHRLHKLQYRTQTISNSLAIDDEAIFRR 45
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 107 VVFWHGFV--GLVGEDAFSMYPNWGTGKYNVKQELFTLGYRVEANVGAFFSSNYDRAVE 164  
 QY 46 -----FRRLDRSLTHASPN-----LVGSHGGLVYK--RY 74  
 Db 165 LYYIKGGRVVDYGAHAHAKYGHKRYGRTYEGIMDMEBKHHLVGHSWGQTLRLMEHF 224  
 QY 75 LES-----RAPSCETLSHVVAIGSPLOGA-----SIV 101  
 Db 225 LKNGGEIDYORQYGVSDLFKGGQDNMVTITTLGTPHNGTPAADKLGSTKFIQDTI 264  
 QY 102 NKIEQLG---LGVALGNSAEFGLEKHDESRYPQKSGSIAGT 140  
 Db 285 NRIGKIGTKALDLELGFSS-OMGFKQOPNES-YAEYAKRIANS 325

## RESULT 11

US-10-282-122A-70475  
 ; Sequence 70475, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282.122A  
 ; PRIOR FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 70475  
 ; LENGTH: 627  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-10-282-122A-70475

Query Match 9.5%; Score 88.5; DB 12; Length 627;  
 Best Local Similarity 21.1%; Pred. No. 1.1;

Matches 47; Conservative 32; Mismatches 55; Indels 89; Gaps 12;

QY 3 IIVHGLMYHGV-----MHP-----LSHRLKLGRTYGTISNSIAID-DEAI- 45  
 Db 248 VVFWHGFV--GLVGEDAFSMYPNWGTGKYNVKQELFTLGYRVEANVGAFFSSNYDRAVE 305

QY 46 -----FRRLDRSLTHASPN-----LVGSHGGLVYK--RY 74  
 Db 306 LYYIKGGRVVDYGAHAHAKYGHKRYGRTYEGIMDMEBKHHLVGHSWGQTLRLMEHF 365  
 QY 75 LES-----RAPSCETLSHVVAIGSPLOGA-----SIV 101  
 Db 366 LKNGGEIDYORQYGVSDLFKGGQDNMVTITTLGTPHNGTPAADKLGSTKFIQDTI 425  
 QY 102 NKIEQLG---LGVALGNSAEFGLEKHDESRYPQKSGSIAGT 140  
 Db 426 NRIGKIGTKALDLELGFSS-OMGFKQOPNES-YAEYAKRIANS 466

## RESULT 12

US-10-156-761-14086  
 ; Sequence 14086, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHITOKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 14086  
 ; LENGTH: 286  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 ; US-10-156-761-14086

Query Match 9.3%; Score 87.5; DB 14; Length 286;  
 Best Local Similarity 23.5%; Pred. No. 0.45;  
 Matches 53; Conservative 23; Mismatches 53; Indels 97; Gaps 12;

QY 3 IIVHGLMYH-----GLVHAPLSRHLKLGRTYGTISNSIA-----IDEA-- 44  
 Db 53 VVFWHGTGNSVDNMLGLAPY-LEHR-----GYCVFSLDYGQSGVFFHGLGPIDKSABQ 107  
 QY 45 --IFRLDRSL--THASPNALVGHSLGLVIRKYLESRAPSCETLSHVVAIGSPLOGASI 100  
 Db 108 LQVF--VDKVLATATGATKADLVGHSQGMPRYTLKF-----LGGACK 148  
 QY 101 VNKIEQLGVALGNSAEFGLEKHDESRYPQKSGSIAGTIFLGRSLRLRPLDSDGTV 160  
 Db 149 VNAL-----VGIANPNH-----GTLISGLNLMLFYPF--GAE 178  
 QY 161 TVEETKIAQMTD-----HIAISTSYEMAY 185  
 Db 179 DLSTATPGIADQVVGSAFMAKLNAGGDTVAGVHTVATATQYDEVV 224

## RESULT 13

US-10-282-122A-47436  
 ; Sequence 47436, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel

```
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
FILE REFERENCE: ELITRA.034A
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 47438
LENGTH: 364
TYPE: PRT
ORGANISM: Burkholderia cepacia
US-10-282-122A-47438

Query Match
Best Local Similarity 9.3%; Score 87.5; DB 12; Length 364;
Matches 34; Conservative 20; Mismatches 43; Indels 27; Gaps 6;

QY 3 IIVHGL-----YHGLVHPLSHRLHKGIRQTTSYSLAID-----EAI 45
DB 55 IIVHGLTGTDXKXAGVLDWYTG-----IQEDLQGHATVYVNLGFSDDSPNGRGQL 109
QY 46 FRRLDRSL--THASPNALVGHSLGIVIKRYLESRAPSCETSHVVAIGSPLOGASIVNK 103
DB 110 LAYVKTIVLANTGATKYNLVGHSGGLT--SRVAAVAP--DIVASVTTTCTPHRSGSEFADF 166

QY 104 IEGL 107
DB 167 VQCV 170

RESULT 14
US-10-282-122A-76642
Sequence 76642, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangshu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 76642
LENGTH: 345
TYPE: PRT
ORGANISM: Treponema pallidum
US-10-282-122A-76642

Query Match
Best Local Similarity 9.3%; Score 87; DB 12; Length 345;
Matches 37; Conservative 26; Mismatches 66; Indels 12; Gaps 4;

QY 45 IFRRLDRSLTHASPNALVGHSLGIVIKRYLESRAPSCETSHVVAIGSPLOGASIVN-- 102
DB 154 LYRVDFTRV-----GLVGHSMGDTALAHY--ANKYP--ERIRTVVLAHPNGADWVNMV 205
QY 103 --KIEQLGLVGLVAGNSAFGLKEDDSRYPOKSGSIAGTIPIGLRSLLRDLSDGTIV 160
DB 206 GKKNEQGLTDLRSVRAVLLITGKQMLAYPERSYEPFQLPDVDVACFLSKMGHGAG 265

QY 161 TVERTKIAGMTDHAISTSY 181
DB 266 TDKYFVWGFTYNTKLSVESF 286

RESULT 15
US-09-938-803-20
Sequence 20, Application US/09938803
Patent No. US2002007672A1
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Tang, Y. Tom
APPLICANT: Lal, Preeti
APPLICANT: Reddy, Roopa
APPLICANT: Baughn, Mariah R.
APPLICANT: Yang, Junming
APPLICANT: Azimzal, Yalda
TITLE OF INVENTION: FULL-LENGTH EXPRESSED GENETIC MARKERS
FILE REFERENCE: PP-0695 US
CURRENT APPLICATION NUMBER: US/09/938,803
CURRENT FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/311,894
PRIOR FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PERL Program
SEQ ID NO 20
LENGTH: 505
TYPE: PRT
ORGANISM: Homo sapiens
```

```

FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone 3039890
US-09-938-803-20

```

```

Query Match          9.3%; Score 87; DB 9; Length 505;
Best Local Similarity 27.0%; Pred. No. 1.2;
Matches 37; Conservative 20; Mismatches 54; Indels 26; Gaps 6;

```

```

Oy 49 LDRSLTHASPVNLYGHSIGLVIRYDESRAPSCETLSHYVAIGSPLOGASIVNKIEQL 107
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 42 LKLMFEEELRNACRGQGVGF-----PAMKOIGNVAAL-----PGIVRSIGLP 86

Oy 108 ----GLGVALGNSAEFGKHEHDESRYPQKSGSIAGTIPGLRSLLRDPLDSGTVVE 163
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 87 DVHSGYGPAIGMAAFDY--NDPEA--VSRFGVGFDINGVR--LIRTNLDESQVQPVK 140

Oy 164 ETKIAGMTDHTAISTTS 180
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 141 EQLAQMFPDHIPIGVGS 157
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Search completed: April 29, 2004, 11:42:30
Job time : 44 secs

```